

=&gt; d que

L4 72 SEA FILE=HCAPLUS ABB=ON PLU=ON ("BARRY C"/AU OR "BARRY C C"/AU OR "BARRY C D"/AU OR "BARRY C DAVID"/AU OR "BARRY C E"/AU OR "BARRY C E III"/AU OR "BARRY C J"/AU OR "BARRY C L"/AU OR "BARRY C M"/AU OR "BARRY C M F"/AU OR "BARRY C MARVIN"/AU OR "BARRY C N"/AU OR "BARRY C P"/AU OR "BARRY C S"/AU OR "BARRY CAROLINE"/AU OR "BARRY CAROLYN B"/AU)

L5 126 SEA FILE=HCAPLUS ABB=ON PLU=ON ("BOUGUELERET L"/AU OR "BOUGUELERET LYDIE"/AU)

L6 130 SEA FILE=HCAPLUS ABB=ON PLU=ON ("CHUMAKOV I"/AU OR "CHUMAKOV I F"/AU OR "CHUMAKOV I M"/AU OR "CHUMAKOV I R"/AU OR "CHUMAKOV I S"/AU OR "CHUMAKOV I V"/AU OR "CHUMAKOV ILYA"/AU OR "CHUMAKOV ILYA M"/AU)

L7 12 SEA FILE=HCAPLUS ABB=ON PLU=ON (L\*\*\* OR L\*\*\* OR L\*\*\*)

L8 27 SEA FILE=HCAPLUS ABB=ON PLU=ON (L4 AND (L5 OR L6 OR L7)) OR (L5 AND (L6 OR L7)) OR (L6 AND L7)

L9 302 SEA FILE=HCAPLUS ABB=ON PLU=ON (L4 OR L5 OR L6 OR L7)

L10 1 SEA FILE=HCAPLUS ABB=ON PLU=ON L9 AND BAP?

L11 23 SEA FILE=HCAPLUS ABB=ON PLU=ON L9 AND PROSTAT?

L12 37 SEA FILE=HCAPLUS ABB=ON PLU=ON L8 OR (L10 OR L11)

=&gt; d ibib abs 1-37

L12 ANSWER 1 OF 37 HCAPLUS COPYRIGHT 2006 ACS on STN

ACCESSION NUMBER: 2006:367166 HCAPLUS

DOCUMENT NUMBER: 144:405904

TITLE: Polynucleotides encoding human TBC-1 protein, polymorphic markers therein, and uses thereof

INVENTOR(S): Blumenfeld, Marta; **Bougueleret, Lydie;**  
**Chumakov, Ilya**

PATENT ASSIGNEE(S): Genset S.A., Fr.

SOURCE: U.S. Pat. Appl. Publ., 114 pp.

CODEN: USXXCO

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
US 2006084073	A1	20060420	US 2004-970750	20041020
PRIORITY APPLN. INFO.:			US 2004-970750	20041020

AB The invention concerns genomic and cDNA sequences of the human TBC-1 gene. The invention also concerns polypeptides encoded by the TBC-1 gene. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the TBC-1 gene useful in genetic anal. The TBC-1 gene is located in a candidate region for **prostate** cancer on chromosome 4. Expression of TBC-1 gene leads to production of at least two mRNA mols. which have different first exons, exon 1 and 1bis, due to alternative splicing.

L12 ANSWER 2 OF 37 HCAPLUS COPYRIGHT 2006 ACS on STN

ACCESSION NUMBER: 2005:182920 HCAPLUS

DOCUMENT NUMBER: 142:258503

TITLE: Secreted polypeptide species in human plasma, detection assays for smaller proteins and tryptic peptides, and expression profiles useful for disease



The invention is directed to BAP28 polypeptides, BAP28 polynucleotide sequences and regulatory region located at the 3' and 5' ends of the BAP28 coding region. The BAP28 polypeptides can be expressed by standard recombinant methodology. BAP28 polynucleotides and polypeptides have been found to be over expressed in prostate tumour cells, therefore levels of BAP28 expression and/or activity may be assayed (e.g. by polymerase chain reaction (PCR)) to diagnose patient suffering from or susceptible to prostate cancer. Antibodies specific for the BAP28 polypeptides are useful as diagnostic reagents. Biallelic markers of the BAP28 gene are useful in genetic analysis. The present sequence represents a protein encoded by a first cDNA sequence of the BAP28 gene consisting of the exons 1 to 45

Sequence 2144 AA;

Query Match	100.0%;	Score 8212;	DB 4;	Length 2144;
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 1629;	Conservative 0;	Gaps 0;		
1	MTSLAQQLRLALPQSDASILSRDEVASLLFDPKKAATIDRTAFAGCTGLEELLGIDP	60		
1	MTSLAQQLRLALPQSDASILSRDEVASLLFDPKKAATIDRTAFAGCTGLEELLGIDP	60		
61	SFEQFEPFLFSQAKTLEERSVQTKAVNKKQDENISLFLHLSPFLKPKQKCLEMLIHR	120		
61	SFEQFEPFLFSQAKTLEERSVQTKAVNKKQDENISLFLHLSPFLKPKQKCLEMLIHR	120		
121	FHHLNODSLIACVLYHEHTRIFVRYIOLKINNSGHRFWLLPVKQSVPLAKGLTIT	180		
121	FHHLNODSLIACVLYHEHTRIFVRYIOLKINNSGHRFWLLPVKQSVPLAKGLTIT	180		
181	HCYKIDGFMDIFCSLYTKSVKVPFAEYPGSSAOLRVLLAFYASTIVSALVAEDVSDNIIA	240		
181	HCYKIDGFMDIFCSLYTKSVKVPFAEYPGSSAOLRVLLAFYASTIVSALVAEDVSDNIIA	240		
241	KLPFYIOGKSKSLPDYRAATYMIICQISVYKMENTFNVSLSAQIKTKITKPSLTKDG	300		
241	KLPFYIOGKSKSLPDYRAATYMIICQISVYKMENTFNVSLSAQIKTKITKPSLTKDG	300		
301	LSCLIYLLORQKPEISGKPPHLCNVPDLITLHGISETYDVSPLRLYMLPHLVSIIH	360		
301	LSCLIYLLORQKPEISGKPPHLCNVPDLITLHGISETYDVSPLRLYMLPHLVSIIH	360		
361	HVTGEETEGMDGOIYKHLBAITKISLKNLDDLASLLPEEYIYSYSGOEMDSNKVSL	420		
361	HVTGEETEGMDGOIYKHLBAITKISLKNLDDLASLLPEEYIYSYSGOEMDSNKVSL	420		
421	LNEQFLPLILLSKYPRTLDVLEHLKEIADLKKQELFHOFVSLSTSGSKYQFLADSD	480		
421	LNEQFLPLILLSKYPRTLDVLEHLKEIADLKKQELFHOFVSLSTSGSKYQFLADSD	480		
481	TSMLSLNHLPLAPYRIILAMNHLKKIMKTSKEGVDESFIKEAVLARLDGDNIDVLSAISA	540		
481	TSMLSLNHLPLAPYRIILAMNHLKKIMKTSKEGVDESFIKEAVLARLDGDNIDVLSAISA	540		
541	FEIFKEHFSSEVTISNLIINFORAELSNQGEWYEVLLIADILLKEELISNDOLSNQV	600		
541	FEIFKEHFSSEVTISNLIINFORAELSNQGEWYEVLLIADILLKEELISNDOLSNQV	600		
601	VCLLPFVIVINDDESAEMKIAIYLSKSGICSLPRLRGFEALENVIKSTPKGLIGVA	660		
601	VCLLPFVIVINDDESAEMKIAIYLSKSGICSLPRLRGFEALENVIKSTPKGLIGVA	660		
661	NQKMIELLADNINIGDPSMLKWEEDLISVGEESFNILKQRTVHVLISLVSCSSLSKE	720		
661	NQKMIELLADNINIGDPSMLKWEEDLISVGEESFNILKQRTVHVLISLVSCSSLSKE	720		
721	THPPPAIRVPSLLQKTKLESVITAVIPESEWIELMDGIVEELMAHVEELNSQR	780		
721	THPPPAIRVPSLLQKTKLESVITAVIPESEWIELMDGIVEELMAHVEELNSQR	780		
781	VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQLEKDSRDYLLHLLIGLFEMNLNGA	840		
781	VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQLEKDSRDYLLHLLIGLFEMNLNGA	840		

Db	781	VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQLEKDSRDYLLHLLIGLFEMNLNGA	840
Qy	841	DAVHFRVLMKFLIKVHLEDPQLFKFCSVLMTYGGSLSNPLNCSVKYVLQTOALYVGCAM	900
Db	841	DAVHFRVLMKFLIKVHLEDPQLFKFCSVLMTYGGSLSNPLNCSVKYVLQTOALYVGCAM	900
Qy	901	ISSQKTOCKHQIASSIPVYTSLLINLSPVKEVRRRAIOCLQALSGVASPRYLLIDHII	960
Db	901	ISSQKTOCKHQIASSIPVYTSLLINLSPVKEVRRRAIOCLQALSGVASPRYLLIDHII	960
Qy	961	SKAEETSDAAVVIDDLATLFEILOREKKLKSQKLSSETLKNLSCVYSCSPYIAKDLMK	1020
Db	961	SKAEETSDAAVVIDDLATLFEILOREKKLKSQKLSSETLKNLSCVYSCSPYIAKDLMK	1020
Qy	1021	VLOGVNGEVLSQLPMAEQLLLEKIQKEPTAVLKDEAVYHLTLGKYNESVSLNEDPK	1080
Db	1021	VLOGVNGEVLSQLPMAEQLLLEKIQKEPTAVLKDEAVYHLTLGKYNESVSLNEDPK	1080
Qy	1081	SDDIFKAVHTTKELYAGPPTIOITALEKITKPPFAAISDEKVQOKLRLMLFDLLVNCN	1140
Db	1081	SDDIFKAVHTTKELYAGPPTIOITALEKITKPPFAAISDEKVQOKLRLMLFDLLVNCN	1140
Qy	1141	SHCAQTVSSVFKGISVNAEQVRIELEPPKAPDLGTVOQKRQKQOKKSODLESVOEVG	1200
Db	1141	SHCAQTVSSVFKGISVNAEQVRIELEPPKAPDLGTVOQKRQKQOKKSODLESVOEVG	1200
Qy	1201	GSYWRVTLILELOHKKRLSPQIIVPTLFWLSKCLEPLPOEGNMETVQOLILSCLL	1260
Db	1201	GSYWRVTLILELOHKKRLSPQIIVPTLFWLSKCLEPLPOEGNMETVQOLILSCLL	1260
Qy	1261	NICQKLSPPGGKIPKQIILDEKPNVELIVQICRLSMPQTHHALLLGTGAGIPDXYL	1320
Db	1261	NICQKLSPPGGKIPKQIILDEKPNVELIVQICRLSMPQTHHALLLGTGAGIPDXYL	1320
Qy	1321	HNIMSIFTEMGANVMLDDTYSFOVINKTVKQVIPALIOSDSDSIEVRNVEIYVKII	1380
Db	1321	HNIMSIFTEMGANVMLDDTYSFOVINKTVKQVIPALIOSDSDSIEVRNVEIYVKII	1380
Qy	1381	SVFVDALPHVPEHRRPLIVQVDTIGAKFLMILLILFEQVYVTVLAAVAGEKDALI	1440
Db	1381	SVFVDALPHVPEHRRPLIVQVDTIGAKFLMILLILFEQVYVTVLAAVAGEKDALI	1440
Qy	1441	EADTEFMFVSCEFSYQHOIQSILMNIQYLKLPKEKEETIPKAVSPNKSQOEMLOVF	1500
Db	1441	EADTEFMFVSCEFSYQHOIQSILMNIQYLKLPKEKEETIPKAVSPNKSQOEMLOVF	1500
Qy	1501	NVETHTSKQLRHFKLVSFMSQILSSNNFLKKVYESGGEPIIKGLERLLETVLGYSIA	1560
Db	1501	NVETHTSKQLRHFKLVSFMSQILSSNNFLKKVYESGGEPIIKGLERLLETVLGYSIA	1560
Qy	1561	VASMERANDKLTVKRWRALLSKAYLLDKVNALLPTEPIPIVIRGVGNPLPSVRKAL	1620
Db	1561	VASMERANDKLTVKRWRALLSKAYLLDKVNALLPTEPIPIVIRGVGNPLPSVRKAL	1620
Qy	1621	DLINNKLLQ 1629	
Db	1621	DLINNKLLQ 1629	

RESULT 2  
ADQ89800  
ID ADQ89800 standard; protein; 2144 AA.  
AC ADQ89800;  
XX  
XX 21-OCT-2004 (first entry)  
XX  
XX Antagonist of cell cycle progression polypeptide #115.  
XX  
XX Cytostatic; cancer; cell division cycle; mitosis; meiosis;  
XX  
XX cell cycle progression.  
XX  
XX Homo sapiens.

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:42:00 ; Search time 39.2894 Seconds  
(without alignments)  
3989.295 Million cell updates/sec

Title: us-09-603-665-5\_COPY\_1\_1629  
Perfect score: 8212  
Sequence: 1 MTSLAQQLRLALPQSDASL.....NPLPSVRRKALDILNNKLLQ 1629

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	673.5	8.2	1649	T39938	hypothetical prote
2	654	8.0	1769	S53378	probable membrane
3	536	6.5	1650	T27864	hypothetical prote
4	248	3.0	2670	T37919	GCN1 homolog - fis
5	241	2.9	2895	H85362	hypothetical prote
6	229	2.8	2059	T41933	large tegument pro
7	228	2.8	3433	S28381	urothrin - human
8	228	2.8	4131	T21085	hypothetical prote
9	227.5	2.8	2954	T14156	kinesin-related pr
10	226	2.8	1447	F82909	hypothetical prote
11	226	2.8	1830	E82969	conserved hypochet
12	224	2.7	2297	AB2494	hypothetical prote
13	219.5	2.7	2819	A90551	conserved hypochet
14	215	2.6	2663	S28261	centromere protein
15	212.5	2.6	4385	T29042	hypothetical prote
16	212	2.6	2108	S28417	CDC39 protein - ye
17	210.5	2.6	2268	T28677	thoptry protein -
18	209	2.5	2401	T28676	thoptry protein -
19	208	2.5	4717	T41581	hypothetical colle
20	207.5	2.5	1655	S47446	nucleoporin Np188
21	207.5	2.5	4859	S74173	ryanodine receptor
22	207	2.5	3744	S46715	hypothetical prote
23	206	2.5	1919	T40032	hypothetical prote
24	204	2.5	3187	JC5837	364k Golgi complex
25	202.5	2.5	2493	S45734	probable membrane
26	202	2.5	3079	RGBY12	probable GTPase-ac
27	201.5	2.5	2672	A48126	translational activa
28	201	2.4	1957	T38077	hypothetical colle
29	199.5	2.4	1964	A59282	nonmuscle myosin I

30	198.5	2.4	1937	2	T41023	probable nuclear p
31	198.5	2.4	3795	2	T00831	hypothetical prote
32	198	2.4	2453	2	T40540	hypothetical prote
33	197.5	2.4	1922	2	T00637	hypothetical prote
34	197.5	2.4	2094	2	S33124	cpf protein - huma
35	197.5	2.4	2166	2	G70163	hypothetical prote
36	196	2.4	1690	2	T13030	microtubule bindin
37	195.5	2.4	1420	2	S54471	probable membrane
38	194.5	2.4	2376	2	S48405	probable membrane
39	194.5	2.4	4872	2	S27272	ryanodine receptor
40	194.5	2.4	4924	2	T50176	probable peptide s
41	194	2.4	2110	2	H96803	unknown protein T5
42	192	2.3	2470	2	S57085	1-phosphatidylinos
43	192	2.3	2712	2	T05113	hypothetical prote
44	191.5	2.3	1732	2	G84664	hypothetical prote
45	191.5	2.3	3655	2	T38084	TRAP-like protein

ALIGNMENTS

RESULT 1									
T39938									
hypothetical protein SPBC2386.04c - fission yeast (Schizosaccharomyces pombe)									
C:Species: Schizosaccharomyces pombe									
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004									
C:Accession: T39938									
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Pohl, T.									
submitted to the EMBL Data Library, May 1998									
A:Reference number: Z21892									
A:Accession: T39938									
A:Status: preliminary; translated from GB/EMBL/DBJ									
A:Molecule type: DNA									
A:Residues: 1-1649 <MOO>									
A:Cross-references: UNIPROT:O60179; UNIPARC:UPI000013AF93; EMBL:AL023287; PDB:CAA18872.1									
A:Experimental source: strain 972h-; cosmid c2386									
C:Genetics:									
A:Gene: SPDB:SPBC2386.04c									
A:Map position: 2									
Query Match									
Best Local Similarity 20.5%; Pred. No. 1.7e-25;									
Matches 348; Conservative 263; Mismatches 490; Indels 593; Gaps 61;									
QY	2	TSLSAQQLRLALPQSDASLSRDEV---ASLLFDPKREAAITIDDTAFALIGCTGLEELGI	58						
DB	3	SSLSQKQKLNKNI---QSN-NVLKINKIRRAPSLLDPKVAADMDEEIVTAVSGHELAHV	58						
QY	59	DPSPEQEPAPLFSQAKTLERSVQTKAVNKQDENISLFLHSPYFLKPAQCKLEMLI	118						
DB	59	EPRLILFEKTLLEBOSVQVRVLLNRTEKIDECQIIRLAPFTEGNALKVLEMLI	118						
QY	119	HRFHILYNODSLIACVLPHEHTRIFPVAVIQLKINKSKRMFWLDPVKGSPVLAGTIL	178						
DB	119	RRRSIHETVSDERFLSPFPHDHPFARILGCSK-PKSRPDLLENAIKMP-VLISRADI	176						
QY	179	ITHCYKDLGFMADICSLVTSKVFAYFPGSSAQLRVLLAFYSTIYSALVAADVDN-	237						
DB	177	VHLSRDKPEFPAPFQVNTASHNNYP-----ELAFWAGTMEVVAHMSNEDP	229						
QY	238	--IIAKLFPYIQGLK--SSLPYRAATWMIICQISVKVMENTFVSLASQILKTLTKI	293						
DB	230	NVLDRFLKAVSAVSVSSI-DPQIAGFMILSSIASLSPSIIIPVLSAIDRLSP-	287						
QY	294	PSLTKDGLSCLIVL-----QRKPESLG-----	317						
DB	288	DNMKPALICVGHILQFCSSFEEDHEQLEKLESFGASSLIEISQERLDEFFVSYVSL	346						
QY	318	-----KKPFPHLCNVPLRLITLHGISETYDVSPPLRLPHLVVSIHHVTGETEGMDG	372						
DB	347	IKSRQKQKRRLLSLSD--TSISQIRVTHQAKFLSVIP-----VNQDFKAL--	392						
QY	373	QIVKRLHAILITKISLKNLIDHLASLLPFEYISYSSQSEMDSNKVSLLNEQFLPLIRLL	432						

[illegible]

```

Db      999 HFCCLITVQSFVARIGISINOCNFRCLKSLBEOGNSDNGKAVSLIKLDELPE-----MD 1053
Oy      1502 VETHTSKOLRHFKEFLSVSMQOLLSSNNF---LKRVESGGEPILKGLEERLLLETVLGYI 1558
Db      1054 VDLATLGSLR-KVLEL--ISLVSKAKNFAPLAKIMENSVSVE-ICAGLPESIKILLI 1109
Oy      1559 SAVAGS---MEKNMCKLTVPKFPARALLSKAYDDLKDKNALLPTEFTPIVRGLGNPLPSV 1615
Db      1110 TLSQQGSNEME-----LGHVVVALRSYIHLLPNELFCTVLGKLHDERALL 1155
Oy      1616 RKKALDLLNNKLOQ 1629
Db      1156 RKALSTIVQORVOQ 1169

RESULT 2
S53378
probable membrane protein YUL109c - yeast (Saccharomyces cerevisiae)
N.Alternate names: hypothetical protein J0808
C.Species: Saccharomyces cerevisiae
C.Date: 05-May-1995 #sequence revision 01-Sep-1995 #text_change 09-Jul-2004
C.Accession: S53378; S56887; S57359
R.Rasmussen, S.W.
submitted to the EMBL Data Library, February 1995
A.Description: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and C
A.Reference number: S53376
A.Accession: S53378
A.Molecule type: DNA
A.Residues: 1-1769 <RAM>
A.Cross-references: UNIPROT:P42945; UNIPARC:UPI0000052F40; EMBL:X65021; NID:g728698; PID:
R.Rasmussen, S.W.
submitted to the Protein Sequence Database, September 1995
A.Reference number: S56876
A.Accession: S56887
A.Molecule type: DNA
A.Residues: 1-1769 <RAM>
A.Cross-references: UNIPARC:UPI0000052F40; EMBL:Z49384; NID:g1008292; PID:g1008293; MIPS
R.Rasmussen, S.W.
Yeast 11, 873-883, 1995
A.Title: A 37.5 kb region of yeast chromosome X includes the SME1, MEF2, GSH1 and CSD3 ge
A.Reference number: S57357; MID:96090136; PMID:7483851
A.Accession: S57359
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-1769 <RAM>
A.Cross-references: UNIPARC:UPI0000052F40; EMBL:X65021; NID:g728698; PID:CMA59385.1; PIR
A.NCBI: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
C.Genetics:
A.Cross-references: SGD:S0003645
A.Map position: 10L
C.Keywords: transmembrane protein

Query Match          8.0%; Score 654; DB 2; Length 1769;
Best Local Similarity 20.8%; Pred.No.1.7e-24;
Matches 366; Conservative 252; Mismatches 542; Indels 598; Gaps 65;

Oy      1 MTSLAQQLRLALPGSDASL-----LSRDENVSLPPPKCAATIDRDTAFALIGCTGLEELL 56
Db      1 MSSLSDOALAQAASNNAVTALDRKRROKTLHASLIYNASKTAATDYDFIFENASKALEBELS 60
Oy      57 GIDPSEFOFEAPLFSQLAKTTERSVOYTKAVNKOLDENISLFTLHSPPYLKPDAQCLEW 116
Db      61 QIEPFALFSRTLPESSSISLDNRNVOTKEEKIDLDVAINAVALLASKWYALPTLNATEW 120
Oy      117 LIHRPHILHNODSLIACVLPYHETRIFPVRIOLKINNKKRWFWLLP-----VK 167
Db      121 LVRRQHIVKNTKEMLLSTLNTNYQHPVFKEIIISTIK-----LPPLFNCLSNFVR 169
Oy      168 QSGVPLAGTLTHCYCKDLGFNDPICSLTVTKSVKVFAEYPGSSAOQLRVLLAFYASTIVSA 227
Db      170 SEKPPTAL-TMI-KLFNMDMLFKLYTSYLDQCICKHNAVYTNQ-----LLFTCCFIN- 219
Oy      228 LVAEDVSNDNIIAKLPFI----QKGKASSLPDYRAATYMTIIQGISVKTMENTFVNLSLA 283

```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:41:34 ; Search time 166.656 Seconds  
(without alignments)  
9041.668 Million cell updates/sec

Title: US-09-603-665-5\_COPY\_1\_1629  
Perfect score: 8212  
Sequence: 1 MTSIAQQLRRLALPQSDASL.....NPLPSYRRAKLDLNNKLOQ 1629

Scoring table: BLOSUM62  
Gapop 10.0 ; Gapext 0.5

Searched: 2849598 seqs, 925015592 residues  
Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 7.2:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8204	99.9	2144	1 HEATR_HUMAN	09h83 homo sapien
2	7755.5	94.4	2063	05T307_HUMAN	05t307 homo sapien
3	4891	59.6	1180	03JUNW7_MOUSE	03jun7 mus musculu
4	4516	55.0	1090	03TNEI1_MOUSE	03tnei1 mus musculu
5	4118	50.1	2159	07SY48_BRARE	07sy48 mus musculu
6	4072	49.6	984	03TWK0_MOUSE	03twk0 mus musculu
7	3589	43.7	2288	04T723_TETNG	04t723 tetradon n
8	3099	37.7	733	08BLJ4_MOUSE	08blj4 mus musculu
9	3085	37.7	733	03T910_MOUSE	03t910 mus musculu
10	2973	36.2	1106	06P197_HUMAN	06p197 homo sapien
11	2298	28.0	1378	07T153_BRARE	07t153 brachydanio
12	2288	28.0	1336	07T152_BRARE	07t152 brachydanio
13	2181.5	26.6	958	1 HEATR_MACFA	09gm44 macaca fasc
14	1916	23.3	897	08NTL7_HUMAN	08ntl7 homo sapien
15	1444	17.6	2104	07PND6_ANGCA	07pnd6 anopheles g
16	1398.5	17.0	2096	06AWS0_DROME	06aws0 drosophila
17	1396.5	17.0	2096	1 HEATR_DROME	09wm75 drosophila
18	948.5	11.6	743	03VIX6_MOUSE	03vix6 mus musculu
19	852	10.4	168	02K1F9_BOVIN	02k1f9 bos taurus
20	832	10.1	2251	04P937_USUMA	04p937 usitlago ma
21	809.5	9.9	1690	08T9E7_DROME	08t9e7 drosophila
22	799.5	9.7	2237	054HL4_DICDI	054hl4 dirosophila
23	786.5	9.6	1830	1 HEATR_ARARH	09c824 aradictopsia
24	774	9.4	2021	05XK02_CRVNE	05xk02 cryptococcu
25	767	9.3	2021	05SNB7_CRVNE	05snb7 cryptococcu
26	763.5	9.3	1857	06BXO6_DEBHA	06bxo6 debryomyc
27	758.5	9.2	1801	05CAFE8_EMENI	05caf8 aspergillus
28	757.5	9.2	2122	2 O5CAF8_ORISA	05caf8 aspergillus
29	715.5	8.7	1802	2 O2ULC6_ASPOR	02ulc6 aspergillus
30	707.5	8.6	1770	2 O6F793_CANGA	06f793 candida gla
31	707.5	8.6	1798	2 O4WLI9_ASPFU	04wli9 aspergillus

## ALIGNMENTS

RESULT 1	HEATR_HUMAN	STANDARD	PRT	2144 AA.
ID	HEATR_HUMAN	09H583; 05T308; 09NM23;		
AC	09H583; 05T308; 09NM23;			
DT	01-JUN-2001	Integrated into UniProtKB/Swiss-Prot.		
DT	19-JUL-2005	sequence version 3.		
DT	07-MAR-2006	entry version 33.		
DE	HEAT repeat-containing protein 1 (Protein BAP28).			
GN	Name=HEATR1; Synonyms=BAP28;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	(1)			
RP	NUCLEOTIDE SEQUENCE AND VARIANTS-SEP-1694; ALA-1854; ASP-1967 AND			
RG	Human chromosome 1 international sequencing consortium;			
RL	Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.			
RN	(2)			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RG	Human chromosome 1 international sequencing consortium;			
RL	Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.			
RN	(3)			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 1777-2144.			
RX	Published=14702039; DOI=10.1038/ng1285;			
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,			
RA	Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,			
RA	Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,			
RA	Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,			
RA	Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Sugawara M.,			
RA	Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,			
RA	Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,			
RA	Niimiya K., Ishihashi T., Yamashita H., Murakawa K., Fujimori K.,			
RA	Tani H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S.,			
RA	Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,			
RA	Kanehori K., Takahashi-Fuji A., Hara H., Tanase T.-O., Nomura Y.,			
RA	Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,			
RA	Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,			
RA	Moriya S., Montiyama H., Satoh N., Takami S., Teraishi Y., Suzuki O.,			
RA	Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,			
RA	Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,			
RA	Yamaguchi M., Watanabe K., Ikegami A., Itakura S., Fukuzumi Y.,			
RA	Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,			
RA	Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,			
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,			
RA	Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senda T.,			
RA	Matsunuma K., Nakajima Y., Mizuno T., Morinaga M., Komatsu T.,			
RA	Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,			
RA	Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,			
RA	Okumura K., Nagase T., Nomura N., Kikuchi H., Masuno Y., Yamashita R.,			

32	700	8.5	1788	2	07R2M6_NEUCR	07r2m6 neurospora
33	696.5	8.5	1774	2	06CJ57_KLUIA	06cj57 kluyveromyc
34	684.5	8.3	1818	2	0592X6_CANAL	0592x6 candida alb
35	682	8.3	1774	2	0754J8_ASHGO	0754j8 ashyia goss
36	673.5	8.2	1649	1	YG44_SCHPO	0601y9 schizosacch
37	654	8.0	1769	1	UTPI0_YEAST	P42945 saccharomyc
38	641	7.8	1635	2	06C457_YARLI	06c457 yarrowia li
39	629.5	7.7	1806	2	041696_GIBZE	041696 gibberella
40	600	7.3	120	2	06P664_HUMAN	06p664 homo sapien
41	556.5	6.8	1660	2	061Y68_CAEER	061y68 caenorhabdi
42	536	6.5	1650	1	HEATR_CAEER	023495 caenorhabdi
43	405.5	4.9	2034	2	0519S9_ENTHI	0519s9 entamoeba h
44	334	4.1	2606	2	04DAC4_TRYCR	04dac4 trypanosoma
45	330.5	4.0	2631	2	038FN3_YITRP	038fn3 trypanosoma

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RL cDNAs.";  
RL Nat. Genet. 36:40-45(2004).  
RN  
RP IDENTIFICATION BY MASS SPECTROMETRY AND SUBCELLULAR LOCATION.  
RX MEDLINE=22317277; PubMed=12429849; DOI=10.1091/mbc.E02-05-0271.  
RA Scherl A., Couté Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,  
Greco A., Hochstrasser D.F., Diaz J.-J.,  
RT "Functional proteomic analysis of human nucleolus.";  
RL Mol. Biol. Cell 13:4100-4109(2002).  
CC  
CC - INTERACTION:  
CC O86M7:XRNI; NDExpel; IntAct=EBI-373098, EBI-372406;  
CC - SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC - SIMILARITY: Belongs to the HEAT1 family.  
CC - SIMILARITY: Contains 1 HEAT repeat.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC  
CC EMBL; AX067150; CAC26776.1; -; Unassigned DNA.  
CC EMBL; AL359921; CA13775.1; -; Genomic DNA.  
CC EMBL; AK001221; BAA91564.1; ALT\_INIT; mRNA.  
CC IntAct: O9H583;  
CC SWISS-2DPAGE; O9H583; HUMAN.  
CC Ensembl; ENSG00000119285; Homo sapiens.  
CC HGNC; HGNC:25517; HEAT1.  
CC GO; GO:0005515; F:protein binding; IPI.  
CC InterPro; IPR012954; BP28 C.  
CC InterPro; IPR000357; HEAT.  
CC Pfam; PF08146; BP28CT; 1.  
CC Pfam; PF02985; HEAT; 4.  
CC PROSITE; PSS0077; HEAT\_REPEAT; FALSE\_NEG.  
CC Nuclear protein; Polymorphism.  
CC CHAIN 1 2144 HEAT repeat-containing protein 1.  
FT REPEAT 2106 2142 HEAT.  
FT VARIANT 1694 1694 N->S.  
FT VARIANT 1854 1854 V->A.  
FT VARIANT 1967 1967 N->D.  
FT VARIANT 2017 2017 E->G.  
FT CONFLICT 348 348 H->R (in Ref. 1).  
FT CONFLICT 607 607 M->V (in Ref. 1).  
SQ SEQUENCE 2144 AA; 242370 MW; 7590EBA17A5F39D CRC64;  
Query Match 99.9%; Score 8204; DB 1; Length 2144;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1627; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 301 LSCILYLVRQKEESGKKPFPPLCNVPLDITLHGISTEYDVSPILRLVPLHYVSIH 360  
DB 301 LSCILYLVRQKEESGKKPFPPLCNVPLDITLHGISTEYDVSPILRLVPLHYVSIH 360  
QY 361 HVTGEETEGMDGQIYRHRFAITTKISLRKNDHLASLLFEETISYSSGOEEDSNKVS 420  
DB 361 HVTGEETEGMDGQIYRHRFAITTKISLRKNDHLASLLFEETISYSSGOEEDSNKVS 420  
QY 421 LNEQPLPLRLLESKYPRTLDDVLEELKEIADLKQELPHQFVSLSTSGKYOPLADSD 480  
DB 421 LNEQPLPLRLLESKYPRTLDDVLEELKEIADLKQELPHQFVSLSTSGKYOPLADSD 480  
QY 481 TSLMLSLNPLAVRLLANMLKKIKTSKEGSDSEFTEAVLARIGDDNDIVLAISA 540  
DB 481 TSLMLSLNPLAVRLLANMLKKIKTSKEGSDSEFTEAVLARIGDDNDIVLAISA 540  
QY 541 FEIFKEHPSSEVITSLNLFQRAELSKNGEWEVLIADILIKEEILSENDQLSNOV 600  
DB 541 FEIFKEHPSSEVITSLNLFQRAELSKNGEWEVLIADILIKEEILSENDQLSNOV 600  
QY 601 VCLLPVAVINDDTESAEMKIAIYLSKSGICSLHPLRGHEALENVIKSTRGKLIQVA 660  
DB 601 VCLLPVAVINDDTESAEMKIAIYLSKSGICSLHPLRGHEALENVIKSTRGKLIQVA 660  
QY 661 NOKMIELADNINLGDPSMLKVEDLISGEESENLKOKVPHYILSVSCSSLKE 720  
DB 661 NOKMIELADNINLGDPSMLKVEDLISGEESENLKOKVPHYILSVSCSSLKE 720  
QY 721 THFPFAIRVFSLLQKKIKKLESYITAVEIPEWHIEIMDRGIPVELMAHYEELNSTOR 780  
DB 721 THFPFAIRVFSLLQKKIKKLESYITAVEIPEWHIEIMDRGIPVELMAHYEELNSTOR 780  
QY 781 VAVEDSVPLVFSIKKTIYALKAKSPKGDIMWNPOLKEDSDYHLILGLEMMINGA 840  
DB 781 VAVEDSVPLVFSIKKTIYALKAKSPKGDIMWNPOLKEDSDYHLILGLEMMINGA 840  
QY 841 DAVHFRVLMKLFKXHEDEVOLFKECSVLMYTGSSLSNPLNCSVKTIVOTQALYYGCAM 900  
DB 841 DAVHFRVLMKLFKXHEDEVOLFKECSVLMYTGSSLSNPLNCSVKTIVOTQALYYGCAM 900  
QY 901 LSSQKQCKHQQLASISSPVVTSLLNLGSPVKEVRRAIQCLOALSGVASPFYLIIDHLI 960  
DB 901 LSSQKQCKHQQLASISSPVVTSLLNLGSPVKEVRRAIQCLOALSGVASPFYLIIDHLI 960  
QY 961 SKAEBITSDAAYVITODLATFPELOREKCLKSHOKLSETLKNLSCVYSPSYIADIMK 1020  
DB 961 SKAEBITSDAAYVITODLATFPELOREKCLKSHOKLSETLKNLSCVYSPSYIADIMK 1020  
QY 1021 VLOGVNGEWSLQPLMAEQLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080  
DB 1021 VLOGVNGEWSLQPLMAEQLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080  
QY 1081 SLDIFIKAVHTTKELVAGMPTIOITALEKITYKPPFAISDEKYOQKILRMLFDLLNCKN 1140  
DB 1081 SLDIFIKAVHTTKELVAGMPTIOITALEKITYKPPFAISDEKYOQKILRMLFDLLNCKN 1140  
QY 1141 SHCAQVTVSSVFKEGISVNAEOVRILEPDPKAPLGIVQKRRQKQOQKSDLESVOEVG 1200  
DB 1141 SHCAQVTVSSVFKEGISVNAEOVRILEPDPKAPLGIVQKRRQKQOQKSDLESVOEVG 1200  
QY 1201 GSYWQVTVLLELLEQHKKRLSPQIIVPTLNLISRLCEPLPEQGNMEYTKOLISCLL 1260  
DB 1201 GSYWQVTVLLELLEQHKKRLSPQIIVPTLNLISRLCEPLPEQGNMEYTKOLISCLL 1260  
QY 1261 NICOGLSPDGGKIPKOILDEKFNVELIVOCIFLSEMPQTHHALLLGVAGIFPDKVL 1320  
DB 1261 NICOGLSPDGGKIPKOILDEKFNVELIVOCIFLSEMPQTHHALLLGVAGIFPDKVL 1320  
QY 1321 HNTMSITFTFGAVWMLDDTYSFOVINKTVKVIYALIOGSDGSDSEVSANVEIIVKII 1380  
DB 1321 HNTMSITFTFGAVWMLDDTYSFOVINKTVKVIYALIOGSDGSDSEVSANVEIIVKII 1380  
QY 1381 SVEVDALPHVBEHRRRLPILVQVLDVTLGAERKELWILLLEFOYVYTKTVLAAVGEKDAL 1440

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:46:09 ; Search time 28.4956 Seconds  
(without alignments)  
5003.834 Million cell updates/sec

Title: US-09-603-665-5\_COPY\_1\_1629  
Perfect score: 8212  
Sequence: 1 MTSLAQOQLRLLAPQSDASL.....NPLPSVRKALDLINKLQQ 1629

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues  
Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /EMC\_Celerra\_SIDS3/prodata/2/iaa/5 COMB pep: \*  
2: /EMC\_Celerra\_SIDS3/prodata/2/iaa/6 COMB pep: \*  
3: /EMC\_Celerra\_SIDS3/prodata/2/iaa/7 COMB pep: \*  
4: /EMC\_Celerra\_SIDS3/prodata/2/iaa/H COMB pep: \*  
5: /EMC\_Celerra\_SIDS3/prodata/2/iaa/PCTUS COMB pep: \*  
6: /EMC\_Celerra\_SIDS3/prodata/2/iaa/RE COMB pep: \*  
7: /EMC\_Celerra\_SIDS3/prodata/2/iaa/backfilest pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	310	3.8	200	US-09-248-796A-15108	Sequence 15108, A
2	228	2.8	3433	US-09-091-501B-10	Sequence 10, Appl
3	228	2.8	3433	US-09-538-092-1136	Sequence 1136, Ap
4	227.5	2.8	2954	US-09-150-867-1	Sequence 1, Appl1
5	227.5	2.8	2954	US-09-724-584-1	Sequence 1, Appl1
6	222.5	2.7	3830	US-09-693-205A-4	Sequence 4, Appl1
7	216.5	2.6	3829	US-09-693-205A-16	Sequence 16, Appl1
8	215	2.6	2663	US-09-538-092-1252	Sequence 1252, Ap
9	214.5	2.6	3829	US-09-693-205A-2	Sequence 2, Appl1
10	212.5	2.6	2662	US-09-595-684B-31	Sequence 31, Appl
11	212	2.6	2108	US-09-538-092-87	Sequence 87, Appl
12	206.5	2.5	186	US-08-973-462-8	Sequence 8, Appl1
13	206.5	2.5	2482	US-08-328-254-6	Sequence 6, Appl1
14	206.5	2.5	348	US-08-353-700-1	Sequence 1, Appl1
15	206.5	2.5	3248	PCT-US95-16216-1	Sequence 1, Appl1
16	203.5	2.5	3210	US-09-538-092-1154	Sequence 1154, Ap
17	202	2.5	3066	US-08-952-127-12	Sequence 12, Appl
18	201.5	2.5	1979	US-09-949-016-6468	Sequence 6468, Ap
19	201.5	2.5	2047	US-09-949-016-7404	Sequence 7404, Ap
20	201.5	2.5	10182	US-09-134-001C-3159	Sequence 3159, Ap
21	198	2.4	3079	PCT-US94-00198-4	Sequence 4, Appl1
22	197.5	2.4	2186	US-09-949-016-10828	Sequence 10828, A
23	197.5	2.4	2349	US-09-538-092-914	Sequence 914, Appl
24	194.5	2.4	2316	US-09-424-783-3	Sequence 3, Appl1
25	194	2.4	2316	US-09-949-016-10280	Sequence 10280, A
26	194	2.4	3056	US-08-508-836A-8	Sequence 8, Appl1

27	194	2.4	3056	1	US-08-629-001A-3	Sequence 3, Appl1
28	194	2.4	3056	1	US-08-874-266-2	Sequence 2, Appl1
29	194	2.4	3056	2	US-08-642-274D-3	Sequence 3, Appl1
30	194	2.4	3056	2	US-08-952-127-3	Sequence 3, Appl1
31	194	2.4	3056	2	US-08-952-014C-3	Sequence 3, Appl1
32	194	2.4	3056	2	US-09-360-416-2	Sequence 2, Appl1
33	194	2.4	3056	2	US-08-984-090-2	Sequence 2, Appl1
34	194	2.4	3056	2	US-10-351-733-1	Sequence 1, Appl1
35	194	2.4	3056	2	US-10-175-225-2	Sequence 2, Appl1
36	193	2.4	2470	2	US-08-265-967C-2	Sequence 2, Appl1
37	193	2.4	2470	2	US-08-305-790B-3	Sequence 3, Appl1
38	192.5	2.3	3057	2	US-09-360-416-3	Sequence 3, Appl1
39	192.5	2.3	3057	2	US-10-175-225-3	Sequence 3, Appl1
40	192	2.3	2474	2	US-08-265-967C-3	Sequence 3, Appl1
41	192	2.3	2474	2	US-08-305-790B-4	Sequence 4, Appl1
42	191	2.3	5024	2	US-09-710-279-2964	Sequence 2964, Ap
43	190.5	2.3	1698	2	US-09-315-793-12	Sequence 12, Appl
44	190	2.3	1093	2	US-09-315-793-52	Sequence 52, Appl
45	190	2.3	1093	2	US-09-538-092-701	Sequence 701, Appl

## ALIGNMENTS

```
RESULT 1
US-09-248-796A-15108
Sequence 15108, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 15108
LENGTH: 200
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-15108

Query Match      3.8%; Score 310; DB 2; Length 200;
Best Local Similarity 38.0%; Pred. No. 7.8e-18;
Matches 76; Conservative 36; Mismatches 70; Indels 18; Gaps 5;

QY      1 MTSLAQOQLR-----ALPQSDASL--SRDEVASLFPKRNATIDRDTAFAGCTGL 52
      |||||
DB      4 MSLSAQQSINEXTASVALDRKRSKLHSR-----SLRDPQQAQDVEYIEINTEGL 59

QY      53 EELLGIDPFEQFEAPLPSQAKTERSVQYAKNKKODENISLPIHSPFLKPAOK 112
      |||||
DB      60 EDLCEDSDFNFKLTLFSETSVNLDNRQTDVLSQDKNDIAFTLVGPGYGLTSSLK 119

QY      113 CLEMILHREHILVYNODSLIACVLPYHETRIFVRYIOLKINNKKHFWMLPVK-OSGV 171
      |||||
DB      120 AEWLVRRHNAHINHAELMILTLALYFQHPVFEVLANIVPKNLQPIFEMLVGYKQDLKT 179

QY      172 PLAKGTLITHCYKDGPMDF 191
      |||||
DB      180 PPASSIL-----KGSAMNF 194

RESULT 2
US-09-091-501B-10
Sequence 10, Application US/09091501B
Patent No. 6518413
GENERAL INFORMATION:
APPLICANT: Tinsley, Jonathon M
```





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:52:34 ; Search time 101.893 Seconds  
(without alignments)  
7405.549 Million cell updates/sec

Title: US-09-603-665-5\_COPY\_1\_1629  
Perfect score: 8212  
Sequence: 1 MTSLAQOLRLALPQSDASL.....NPLPSVRKALDLINKLQQ 1629

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA Main:  
1: /EMC\_Celextra\_SIDS3/prodata/2/pubppa/US07\_PUBCOMB.pep.\*  
2: /EMC\_Celextra\_SIDS3/prodata/2/pubppa/US08\_PUBCOMB.pep.\*  
3: /EMC\_Celextra\_SIDS3/prodata/2/pubppa/US09\_PUBCOMB.pep.\*  
4: /EMC\_Celextra\_SIDS3/prodata/2/pubppa/US10\_PUBCOMB.pep.\*  
5: /EMC\_Celextra\_SIDS3/prodata/2/pubppa/US108\_PUBCOMB.pep.\*  
6: /EMC\_Celextra\_SIDS3/prodata/2/pubppa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8212	100.0	2144	US-10-745-237-230	Sequence 230, App
2	8195	99.8	2144	US-11-124-368A-277	Sequence 277, App
3	7635	93.0	2044	US-11-124-368A-278	Sequence 278, App
4	7627.5	92.9	1569	US-10-450-763-45591	Sequence 45591, A
5	7266	88.5	2036	US-11-124-368A-276	Sequence 276, App
6	7266	88.5	2036	US-11-124-368A-280	Sequence 280, App
7	7266	88.5	2036	US-11-124-368A-281	Sequence 281, App
8	3694	45.0	734	US-10-108-260A-3300	Sequence 3300, App
9	2824.5	34.4	1149	US-10-128-558-167	Sequence 167, App
10	1916	23.3	897	US-10-764-425-177	Sequence 177, App
11	1603	19.5	325	US-10-128-558-358	Sequence 358, App
12	1396.5	17.0	2096	US-10-745-237-56	Sequence 36, App1
13	1396.5	17.0	2096	US-11-097-143-22518	Sequence 22518, A
14	705.5	8.6	1798	US-10-128-714-8388	Sequence 8388, App
15	704.5	8.6	1814	US-10-128-714-3388	Sequence 3388, App
16	684.5	8.3	1818	US-10-032-585-7616	Sequence 7616, App
17	674.5	8.2	2122	US-10-437-963-189782	Sequence 189782, App
18	663	8.1	1335	US-10-264-049-2412	Sequence 2412, App
19	337	4.6	77	US-09-864-761-37036	Sequence 37036, A
20	337	2.9	952	US-10-424-599-248111	Sequence 248111, A
21	228	2.8	3433	US-10-408-765A-731	Sequence 731, App
22	228	2.8	3433	US-10-408-765A-732	Sequence 732, App
23	228	2.8	3433	US-10-756-149-5681	Sequence 5681, App
24	228	2.8	4131	US-10-369-493-5136	Sequence 5136, App
25	227.5	2.8	2954	US-10-650-280-1	Sequence 1, App1
26	225.5	2.7	1583	US-10-408-765A-1635	Sequence 1635, App
27	218.5	2.7	2543	US-10-828-985A-9	Sequence 9, App1

28	218	2.7	2665	US-11-124-368A-214	Sequence 214, App
29	218	2.7	2668	US-11-124-368A-215	Sequence 215, App
30	217.5	2.6	2228	US-10-511-096-2	Sequence 2, App1
31	217.5	2.6	2230	US-10-511-096-4	Sequence 4, App1
32	216	2.6	2250	US-10-511-096-6	Sequence 6, App1
33	216	2.6	2252	US-10-511-096-8	Sequence 8, App1
34	215.5	2.6	2503	US-10-828-985A-11	Sequence 11, App1
35	215	2.6	2568	US-10-828-985A-7	Sequence 7, App1
36	215	2.6	2633	US-10-450-763-36864	Sequence 36864, A
37	215	2.6	2663	US-10-723-860-749	Sequence 749, App
38	214.5	2.6	3829	US-10-408-765A-849	Sequence 849, App
39	214.5	2.6	3829	US-10-723-860-2137	Sequence 2137, App
40	213	2.6	1581	US-10-437-963-196981	Sequence 196981, App
41	212	2.6	2108	US-10-369-493-1537	Sequence 1537, App
42	210.5	2.6	2733	US-11-097-143-9003	Sequence 9003, App
43	206.5	2.5	1786	US-09-742-096-3	Sequence 3, App1
44	206.5	2.5	1786	US-11-196-400-3	Sequence 3, App1
45	206.5	2.5	1787	US-10-415-253-2	Sequence 2, App1

## ALIGNMENTS

RESULT 1					
US-10-745-237-230					
Sequence 230 Application US/10745237					
Publication/No. US20050227301A1					
GENERAL INFORMATION:					
APPLICANT: <del>Cyotec</del> Limited					
APPLICANT: Glover, David					
APPLICANT: Bell, Graham					
APPLICANT: Frezz, Lisa					
APPLICANT: Midgley, Carol					
TITLE OF INVENTION: Cell Cycle Progression Proteins					
FILE REFERENCE: P0158190 CYK					
CURRENT APPLICATION NUMBER: US/10745,237					
CURRENT FILING DATE: 2003-12-23					
PRIOR APPLICATION NUMBER: US 60/439,123					
PRIOR FILING DATE: 2003-01-10					
PRIOR APPLICATION NUMBER: US 60/468,402					
PRIOR FILING DATE: 2003-05-06					
NUMBER OF SEQ ID NOS: 600					
SOFTWARE: Patent version 3.1					
SEQ ID NO 230					
LENGTH: 2144					
TYPE: PRT					
ORGANISM: Homo sapiens					
FEATURE:					
OTHER INFORMATION: Q9H583					
US-10-745-237-230					
Query Match					
Best local Similarity 100.0%; Pred. No. 0;					
Matches 1629; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MTSLAOLQRLALPQSDASLSRDEVASLFDPKKATITRDPAFAIGCTGLEBLIDP	60	Sequence 214, App	
DB	1	MTSLAOLQRLALPQSDASLSRDEVASLFDPKKATITRDPAFAIGCTGLEBLIDP	60	Sequence 215, App	
QY	61	SFQGFAPLPSQAKTLERSVOTKAVNKQJDNISFLHLSPYFLPKPAOKCLEWLIHR	120	Sequence 2, App1	
DB	61	SFQGFAPLPSQAKTLERSVOTKAVNKQJDNISFLHLSPYFLPKPAOKCLEWLIHR	120	Sequence 4, App1	
QY	121	FHHLNVDSLICVLPYHETRFVAVIQLKTNNSGRHFWFLLPVKQSGVPLAKGLIT	180	Sequence 6, App1	
DB	121	FHHLNVDSLICVLPYHETRFVAVIQLKTNNSGRHFWFLLPVKQSGVPLAKGLIT	180	Sequence 8, App1	
QY	181	HCKKDGFMDPISLTVKSVKFAEYFGSSAQRLVLAFASTIVSALVAEDVSDNIIA	240	Sequence 11, App1	
DB	181	HCKKDGFMDPISLTVKSVKFAEYFGSSAQRLVLAFASTIVSALVAEDVSDNIIA	240	Sequence 7, App1	
QY	241	KLPPIYQKGLSKSLPYRAATYMIQIISVKTMENTFVNSLASQIITKLTPLSLIKDG	300	Sequence 36864, A	
DB	241	KLPPIYQKGLSKSLPYRAATYMIQIISVKTMENTFVNSLASQIITKLTPLSLIKDG	300	Sequence 749, App	

```

Db 241 KLFYIQRKGLSSLPDYRAATYMIICQISVKVMENTFVNSLASQIIKTLTKIPSLIKDG 300
Qy 301 LSCILVLLQROKPESELGKKPFPHLGNVDPDLITLHGISETYDVSPDLRYMLPHLVASIIH 360
Db 301 LSCILVLLQROKPESELGKKPFPHLGNVDPDLITLHGISETYDVSPDLRYMLPHLVASIIH 360
Qy 361 HVTGEETGMDGOIYKRLHAILITKISLKNLIDHLLASLLEBYISYSSQOEEMDSNKVSL 420
Db 361 HVTGEETGMDGOIYKRLHAILITKISLKNLIDHLLASLLEBYISYSSQOEEMDSNKVSL 420
Qy 421 LNEOGLPLIRLESKYPRTLDVYLEEHLKEIADLKOELEHQFVSLSTSGGKYQGLASD 480
Db 421 LNEOGLPLIRLESKYPRTLDVYLEEHLKEIADLKOELEHQFVSLSTSGGKYQGLASD 480
Qy 481 TSLMLSLNHPALAPVAILAMNHLKIMKTSKEGVDSFIKEAVLARIAGDNDIVLSAISA 540
Db 481 TSLMLSLNHPALAPVAILAMNHLKIMKTSKEGVDSFIKEAVLARIAGDNDIVLSAISA 540
Qy 541 PEIFKEHFSSEYITISNLINLFORAELSKNGEWEVYLKIAADLILKEEILSENDOLSNQV 600
Db 541 PEIFKEHFSSEYITISNLINLFORAELSKNGEWEVYLKIAADLILKEEILSENDOLSNQV 600
Qy 601 VCLLPFVYINDDPESAEKIAIYLSKSGICSLHPLLGMEBALENVIKSTKPGKLGVA 660
Db 601 VCLLPFVYINDDPESAEKIAIYLSKSGICSLHPLLGMEBALENVIKSTKPGKLGVA 660
Qy 661 NQKMIELADNINLGDPSMLKMEVDLISVGEESFNLKOKVTPHVLISVLVSCSSLIKE 720
Db 661 NQKMIELADNINLGDPSMLKMEVDLISVGEESFNLKOKVTPHVLISVLVSCSSLIKE 720
Qy 721 THPPAIRVSELLOKKIKLESVITAVEIPSEMHIELMDRGIPEVLAHYVEELNSTOR 780
Db 721 THPPAIRVSELLOKKIKLESVITAVEIPSEMHIELMDRGIPEVLAHYVEELNSTOR 780
Qy 781 VAVDSVFLVSLKKEFIYALKAPKSPFGDIIWMNEQEKEDRDYLLHILIGLFEMMLNGA 840
Db 781 VAVDSVFLVSLKKEFIYALKAPKSPFGDIIWMNEQEKEDRDYLLHILIGLFEMMLNGA 840
Qy 841 DAVHFRVLMKLFIVKHLBDFVQLFKFCGLVMTYSSLSNPLNCSVKTVLQTOALVYGAM 900
Db 841 DAVHFRVLMKLFIVKHLBDFVQLFKFCGLVMTYSSLSNPLNCSVKTVLQTOALVYGAM 900
Qy 901 LSSQKTOCKHOLASISSPVVTSLLINLGSVKEVRRALIOCCQALSGVASPYYLIDHLI 960
Db 901 LSSQKTOCKHOLASISSPVVTSLLINLGSVKEVRRALIOCCQALSGVASPYYLIDHLI 960
Qy 961 SKABEITSDAAYVIOIDLATLFEELQREKKLKSQKLSSETLKNLSCVYSCPSYIAKDLMK 1020
Db 961 SKABEITSDAAYVIOIDLATLFEELQREKKLKSQKLSSETLKNLSCVYSCPSYIAKDLMK 1020
Qy 1021 VLOGVNGEMVLSQLPMAEQULLEKIQKEPTAVLKDCAVVLHLTLGKYNFVSVLINDEPK 1080
Db 1021 VLOGVNGEMVLSQLPMAEQULLEKIQKEPTAVLKDCAVVLHLTLGKYNFVSVLINDEPK 1080
Qy 1081 SLDFIKAVHTTKELIYAGMPTQITALEKITKPPRAAISDEKVQOKLRLMFDLLVNNKN 1140
Db 1081 SLDFIKAVHTTKELIYAGMPTQITALEKITKPPRAAISDEKVQOKLRLMFDLLVNNKN 1140
Qy 1141 SHCAQTVSVFSGISVNAEOVRIELEPPDKAKPLGTVOQKRQKQKQKKSODLESVQEVG 1200
Db 1141 SHCAQTVSVFSGISVNAEOVRIELEPPDKAKPLGTVOQKRQKQKQKKSODLESVQEVG 1200
Qy 1201 GSYWQVTVLILLELOHKKKLSPOILVPTLPMILSRCLBPLFOEGGNEITKQILISCLL 1260
Db 1201 GSYWQVTVLILLELOHKKKLSPOILVPTLPMILSRCLBPLFOEGGNEITKQILISCLL 1260
Qy 1261 NICQGLSDGGKIPDIIIDEEKFNVELIYQCIRLSEMOTHHALLLGTAVAGIPDPKVL 1320
Db 1261 NICQGLSDGGKIPDIIIDEEKFNVELIYQCIRLSEMOTHHALLLGTAVAGIPDPKVL 1320
Qy 1321 HNINSIPTFMGAVNRDLDTYSFOVINKTVKVVIPALIQSDSGDSIEVSRNVEEIVKXI 1380
Db 1321 HNINSIPTFMGAVNRDLDTYSFOVINKTVKVVIPALIQSDSGDSIEVSRNVEEIVKXI 1380

```

```

Qy 1381 SVFVDALPHVEHNRRLPILVOLVDTLGAKEFLMILLILPEQVYTKTVLAAAYEKDAL 1440
Db 1381 SVFVDALPHVEHNRRLPILVOLVDTLGAKEFLMILLILPEQVYTKTVLAAAYEKDAL 1440
Qy 1441 EADTFEPMFVSVCESVQHOLOSINIIQYLLKLPBEKEETIPKAVSPKSSQOEMLQVF 1500
Db 1441 EADTFEPMFVSVCESVQHOLOSINIIQYLLKLPBEKEETIPKAVSPKSSQOEMLQVF 1500
Qy 1501 NVEHTSKOLRHFFELSVSFMQSLLSSNNPLKRVESGGPEILKGLBERLLETVLGYISA 1560
Db 1501 NVEHTSKOLRHFFELSVSFMQSLLSSNNPLKRVESGGPEILKGLBERLLETVLGYISA 1560
Qy 1561 VAQSMERNADLTVEFWRALISKAYDLDKVNALLPTEFTIPVIRGLVGNPLPSVRKAL 1620
Db 1561 VAQSMERNADLTVEFWRALISKAYDLDKVNALLPTEFTIPVIRGLVGNPLPSVRKAL 1620
Qy 1621 DLNNKLOQ 1629
Db 1621 DLNNKLOQ 1629

RESULT 2
US-11-124-368A-277
; Sequence 277, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Lake
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE REFERENCE: C0001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; PRIOR APPLICATION NUMBER: US 60/566,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 277
; LENGTH: 2144
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-368A-277

Query Match 99.8%; Score 8195; DB 6; Length 2144;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1626; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MTSLAQOLQRLAPQSDASLISRDVASSLLPDPKAAITIDRTAFAGCTGLBELLGIDP 60
Db 1 MTSLAQOLQRLAPQSDASLISRDVASSLLPDPKAAITIDRTAFAGCTGLBELLGIDP 60
Qy 61 SFEQPEAPLPBQAKTERSVQOTAVNKKOLDENISLFIHISPEFLKPAQKLEMLIHR 120
Db 61 SFEQPEAPLPBQAKTERSVQOTAVNKKOLDENISLFIHISPEFLKPAQKLEMLIHR 120
Qy 121 FHIHLYNQDSLACVLPEYHETRIFVRVYIOLKIKNSKRWFMWLPVYQSGVPLAKGTLLT 180
Db 121 FHIHLYNQDSLACVLPEYHETRIFVRVYIOLKIKNSKRWFMWLPVYQSGVPLAKGTLLT 180
Qy 181 HCYDGLGMDFTGSLVTKSVKVPFAEYRGSSAQQLAVLAFVASTIVSALVAEDVSDNIIA 240
Db 181 HCYDGLGMDFTGSLVTKSVKVPFAEYRGSSAQQLAVLAFVASTIVSALVAEDVSDNIIA 240
Qy 241 KLFPIYIQKGLSSLPDYRAATYMIICQISVKVMENTFVNSLASQIIKTLTKIPSLIKDG 300
Db 241 KLFPIYIQKGLSSLPDYRAATYMIICQISVKVMENTFVNSLASQIIKTLTKIPSLIKDG 300
Qy 301 LSCILVLLQROKPESELGKKPFPHLGNVDPDLITLHGISETYDVSPDLRYMLPHLVASIIH 360
Db 301 LSCILVLLQROKPESELGKKPFPHLGNVDPDLITLHGISETYDVSPDLRYMLPHLVASIIH 360

```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:53:49 ; Search time 9.06679 Seconds  
(without alignments)  
4243.670 Million cell updates/sec

Title: US-09-603-665-5\_COPY\_1\_1629  
Perfect score: 8212  
Sequence: 1 MTSIAQOLRLALPQSDASL.....NPLPSVRKALDLNNKQ 1629

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 103426 seqs, 23619683 residues

Total number of hits satisfying chosen parameters: 103426

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_NEW\_PUB pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEW\_PUB pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_NEW\_PUB pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_NEW\_PUB pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/PCT\_NEW\_PUB pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10\_NEW\_PUB pep.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_NEW\_PUB pep.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3694	45.0	734	US-11-293-697-3300	Sequence 3300, App
2	206.5	2.5	3113	US-10-505-928-325	Sequence 325, App
3	188.5	2.3	2209	US-11-301-554-1903	Sequence 1903, App
4	171	2.1	1325	US-10-449-902-51794	Sequence 51794, A
5	166.5	2.0	1206	US-10-449-902-43305	Sequence 43305, A
6	165	2.0	1176	US-10-449-902-46322	Sequence 46322, A
7	163.5	2.0	1328	US-10-504-973-32	Sequence 32, App
8	159.5	1.9	2671	US-10-505-928-784	Sequence 784, App
9	158.5	1.9	1116	US-10-449-902-47106	Sequence 47106, A
10	157.5	1.9	980	US-11-289-102-296	Sequence 296, App
11	157	1.9	1205	US-11-293-697-2967	Sequence 2967, App
12	155	1.9	764	US-10-449-902-45961	Sequence 45961, A
13	155	1.9	1842	US-10-511-937-2929	Sequence 2929, App
14	152.5	1.9	1230	US-10-511-937-2929	Sequence 378, App
15	151.5	1.8	1218	US-10-449-902-47146	Sequence 47146, A
16	150	1.8	1050	US-10-449-902-4227	Sequence 4227, App
17	147.5	1.8	1443	US-10-471-571A-4624	Sequence 14, App
18	146.5	1.8	1146	US-10-486-020-14	Sequence 14, App
19	146.5	1.7	1056	US-10-449-902-45516	Sequence 45516, A
20	141	1.7	1346	US-10-449-902-55021	Sequence 55021, A
21	140.5	1.7	1663	US-10-505-928-227	Sequence 227, App
22	139.5	1.7	1174	US-10-449-902-49996	Sequence 49996, A
23	139.5	1.7	1346	US-10-449-902-55428	Sequence 55428, A
24	139.5	1.7	2364	US-11-289-102-242	Sequence 242, App
25	136.5	1.7	1575	US-10-505-928-257	Sequence 257, App

26	136	1.7	1315	US-10-527-411-141	Sequence 141, App
27	135.5	1.7	838	US-10-449-902-45252	Sequence 45252, A
28	135.5	1.7	1479	US-10-471-571A-2392	Sequence 2392, App
29	134.5	1.7	763	US-10-449-902-45594	Sequence 45594, A
30	134	1.6	1166	US-10-449-902-37528	Sequence 37528, A
31	133.5	1.6	712	US-11-293-697-3282	Sequence 3282, App
32	133.5	1.6	1085	US-10-505-928-343	Sequence 343, App
33	133.5	1.6	1158	US-10-471-571A-4944	Sequence 4944, App
34	132.5	1.6	784	US-10-449-902-43318	Sequence 43318, A
35	132.5	1.6	926	US-10-449-902-43257	Sequence 43257, A
36	132	1.6	889	US-10-449-902-36405	Sequence 36405, A
37	131.5	1.6	836	US-11-105-233-154	Sequence 154, App
38	131.5	1.6	995	US-10-471-571A-3910	Sequence 3910, App
39	131.5	1.6	1866	US-10-511-937-2968	Sequence 2968, App
40	130.5	1.6	726	US-11-334-351-2	Sequence 2, App
41	130.5	1.6	3176	US-11-105-233-155	Sequence 155, App
42	130	1.6	732	US-11-203-828-7	Sequence 7, App
43	130	1.6	732	US-11-221-332-102	Sequence 102, App
44	130	1.6	882	US-11-122-986-140	Sequence 140, App
45	130	1.6	882	US-11-122-986-142	Sequence 142, App

ALIGNMENTS

RESULT 1  
US-11-293-697-3300  
; Sequence 3300, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: Patentn Ver. 2.1  
; SEQ ID NO 3300  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-293-697-3300

Query Match 45.0%; Score 3694; DB 7; Length 734;  
Best Local Similarity 100.0%; Pred. No. 1.7e-208;  
Matches 734; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTSLAOLRLALPQSDASLSRDEVASLFDPEKATIDRTAFAGTGLEBLGIDP	60
DB	1	MTSLAOLRLALPQSDASLSRDEVASLFDPEKATIDRTAFAGTGLEBLGIDP	60
QY	61	SFQEFAPLPSQAKTLERSVQTKAVNKQIDENISFLHLSPEFLKPKCKLEWLIHR	120
DB	61	SFQEFAPLPSQAKTLERSVQTKAVNKQIDENISFLHLSPEFLKPKCKLEWLIHR	120
QY	121	FHHLNMODSLIACVLPYHETRIPIVRYIOLKINNSGHRFWMLPVKQSGVPLAKGLTIT	180
DB	121	FHHLNMODSLIACVLPYHETRIPIVRYIOLKINNSGHRFWMLPVKQSGVPLAKGLTIT	180
QY	181	HCYKDLGFMDFICSLVTKSVKFAEPYGGSAQRLVLLAFVASTIVSALVAEDVSDNIIA	240
DB	181	HCYKDLGFMDFICSLVTKSVKFAEPYGGSAQRLVLLAFVASTIVSALVAEDVSDNIIA	240
QY	241	KLPFYIOGKGLSSLPDRAATYMIICQISVAVTMENTFVNSLASOIIKTLTKRISLNDG	300
DB	241	KLPFYIOGKGLSSLPDRAATYMIICQISVAVTMENTFVNSLASOIIKTLTKRISLNDG	300
QY	301	LSCILIVLQRPSPSLGKKPPHLCNVPDLITLIIHGISETYDVSPBLRYMLPHLVVSIH	360
DB	301	LSCILIVLQRPSPSLGKKPPHLCNVPDLITLIIHGISETYDVSPBLRYMLPHLVVSIH	360

Qy	361	HTYGEETEGMDQIYKRHLEALITKISLKNNDHLLASLLFEEYISYSQGEEMDSNKVSL	420
Db	361	HTYGEETEGMDQIYKRHLEALITKISLKNNDHLLASLLFEEYISYSQGEEMDSNKVSL	420
Qy	421	LNBOFLPLRLLESXYPRTLDVLEBHLKEIADLKQOELFHQFVSLSTSGXGYOFPLASD	480
Db	421	LNBOFLPLRLLESXYPRTLDVLEBHLKEIADLKQOELFHQFVSLSTSGXGYOFPLASD	480
Qy	481	TSMLSLNHPPLAPVRILANNHLKKIMTKISKEGVDSFIKEAVLARIQDNDIVVLSAISA	540
Db	481	TSMLSLNHPPLAPVRILANNHLKKIMTKISKEGVDSFIKEAVLARIQDNDIVVLSAISA	540
Qy	541	FEIFKEHPSESVTISNLLPQRAELSKNGWEYVUKIADLILKEILISENDOLSNQV	600
Db	541	FEIFKEHPSESVTISNLLPQRAELSKNGWEYVUKIADLILKEILISENDOLSNQV	600
Qy	601	VCLLPFVVVINDDTESAEMKIAIYLSKSGICSLPLRLRQMEALENVIKSTPKGLIGVA	660
Db	601	VCLLPFVVVINDDTESAEMKIAIYLSKSGICSLPLRLRQMEALENVIKSTPKGLIGVA	660
Qy	661	NQKMTIELADNINLGDPSMSLKMVEDLISVGEESFNLKQKVTFHVILSVLVS CCSLKE	720
Db	661	NQKMTIELADNINLGDPSMSLKMVEDLISVGEESFNLKQKVTFHVILSVLVS CCSLKE	720
Qy	721	THFPFAIRVFSLQ 734	
Db	721	THFPFAIRVFSLQ 734	
RESULT 2			
US-10-505-928-325			
; Sequence 325, Application US/10505928			
; Publication No. US20060088532A1			
; GENERAL INFORMATION:			
; APPLICANT: Ludwig Institute for Cancer Research et al.			
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES			
; FILE REFERENCE: 28967/39178			
; CURRENT APPLICATION NUMBER: US/10/505,928			
; CURRENT FILING DATE: 2004-08-27			
; PRIOR APPLICATION NUMBER: US 60/363,019			
; PRIOR FILING DATE: 2002-03-07			
; NUMBER OF SEQ ID NOS: 866			
; SOFTWARE: PatentIn 3.2			
; SEQ ID NO 325			
; LENGTH: 3113			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-505-928-325			

[illegible]

QY	529	UNIDVLSAISAFIEFKHESSEPT-----ISNLMLFORALSKNGEYVULKIAAD	561
Db	1658	DTEBAIOGNEESCISKSH-TSETTEBTPKHDVHOICDQAQOOLN-----LDIEKTET	1711
QY	582	ILIE--EISENDOLSONOVVCLLPVIVINNDTESAEMKIAIYLSKSGICSHPL-LR	638
Db	1712	GALPTEGCEGSEQSDPN-----YEPGEDTQSSSEIS-ELFSPGNALVPMDFL	1762
QY	639	GWEBALENV--IKSTYPGKL-----IGVANOQKIELIAD-----NINADPSMLKM-	683
Db	1763	GNQEDIHNLQURVETENENMLRLHVIEDDRKVESLLNEKEDSKLHHQEOVLMKIE	1822
QY	684	----VEDLISGEESEFNLOKVTF-----HYLSVLVSCSSSLK---ETHPPAIRV---	729
Db	1823	ACTELEKTIYVGLKKNESDLSSEKLEYFESCDHOELLORVETSEGLNSDLEPMADKSRDIG	1882
QY	730	-----FSLLOKKIKKLESITVAVEIPESEWHIELMDRGIPEVIMAHYVELN	776
Db	1883	DNVAKVNDSWKERPLDVENEELSRISSEKASIE-----HEALYLEADL-----EVV	1927
QY	777	STQVAV-----EDSVLVPFSLKKFIYALKA PKSPKGD-----MNPNOLKEDSRDYL	826
Db	1928	QTEVLCLEKONENQOKVITCLIEELSVAVTSBRNDREBLTMSKTTALDOLSEKMEKT	1967
QY	827	HLIIGLFEMMLNGADAVHFRVLMCLPIKYHLEDVFOLEKFCPSVMTYSSLSNPLNCSVK	866
Db	1988	Q-----ELSHQSECHICIQVAEAEVKETE-----LLQTLSDVSSELKDKHT	2031
QY	867	TVLQTOALVYGCAMLSQKTOCKHOLA-----SISSP	918
Db	2032	LOEKLQSLERDSQALSTTCKELENOIQLNKEKELLVKESESIQARLSSEDYEXLNVSKA	2091
QY	919	VVTSLL-----INLGSVPKEVR--RAAIOLOLGSVAPFYLLIDHLSKAEITSDA	970
Db	2092	LEALVKEGFALRLSTQREHIVQLRGCIKLRIRIADKKOL--HIANKLERREN	2148
QY	971	AYVIOIATTFEELOREKTL-KSHQKL-----SETLKNLSCVYSCPYIAKDL	1018
Db	2149	-----DSLKDKVENLERELQMSBENOELVILDAENSKAEVELTKQIEMARSLKVFELD	2204
QY	1019	M-----KVLGVNGEM-VLSOLPYMAEOLEKIOKEPTAV-LKDEAM-----VLHL	1062
Db	2205	VTLRSEKENTKQIQOKQOQULSELDKLSSPKSILE--EKQOAEIQIKESKTAVENTON	2262
QY	1063	TLQKYNESVSLNED-----PKSIDIFIKAVHTTKELVAGMPTIOITALEKTIKCPFA	1116
Db	2263	QKELNEVAVALCGDQOIMKATEQSDLPPIEEHQUR-----NSIEIKRLALEA	2311
QY	1117	AISBEKVOQLKRLPLVLYNCKNSHCAQVSSVFKIS-----VNAQOVRELEPPD	1169
Db	2312	---DEKOLCVLQOLKE---SEHNADLLKGRVENLELEIARTNOEHAALAEANSK	2362
QY	1170	-----KAPLGTVQOKRQOMOOKSODLESVOEYSGSVWORVTLILIEDLOHKKLRSP	1223
Db	2363	GEVETLAKAKIEGMOISARGLEL-----DVVITRS-----EKENLNLNLOKEGERISEL	2410
QY	1224	QIULVPTLFNILSRCLERPEOQGNMEYTKOLILSCLNLNOOKSLPDGSKIPKD--ILDEE	1281
Db	2411	EIINSSPENI-----LOEKEOEKQOMKEKSSSTAMEMLOTOULKEINERVAALHNDQEACAK	2466
QY	1282	KFNVELLVOCIRLSE-----MPQTHHALLLLGTAVGIF-----PRKVLANI	1323
Db	2467	KONLSQVECELEKALLOGLDBEAKKNYVILQSSVNGLLQEVEDGQOKYLEKQDEBISRL	2526
QY	1324	MSIPTFGANVMRLDDT-----YSFOVINKTYVMVIP-ALIOGSDG--DSIEV--	1368
Db	2527	KNQIODOEOALVSKSYVEGEOHJWKEONLEIRLIVTALJEOKIQVLOSKNSALDPTLEVLQ	2586
QY	1369	--SNNVE--EIVVKISVFDVLLPNHPE-----HRRPITVOLVPTLGAEK--FL	1412
Db	2587	SSYNNLENLELTPMDKMSFVEKYNKNTAKETELOREMHMEAOKTAELQOELSCEKRLA	2646
QY	1413	WILLIILFEQVVTKTVAALVYGEKDALLEADTEFWFVCC-----	1452

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:41:09 ; Search time 144.903 Seconds  
(without alignments)

6765.010 Million cell updates/sec

Title: US-09-603-665-5  
10803

Sequence: 1 MTSLLAQQLQLALPQSDASL.....CQKTIQQLETVLGEPLOSIF 2144

Scoring table: BIOSUM62  
Gapop 10.0 , Capext 0.5

Searched: 2589679 seqs, 45726429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq.8:\*  
1: geneseqp1980s:\*  
2: geneseqp1980s:\*  
3: geneseqp2000s:\*  
4: geneseqp2000s:\*  
5: geneseqp2000s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003as:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*  
10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	10799	99.9	2144	4	AAB85029
2	10799	99.9	2144	8	ADQ89800
3	10791	99.9	2144	8	AEEO2859
4	10782	99.8	2144	8	ADSD34902
5	10222	94.6	2044	8	ADSD34907
6	9853	91.2	2036	8	ADSD34905
7	9853	91.2	2036	8	ADSD34903
8	9853	91.2	2036	8	ADSD34904
9	7756.5	71.8	1569	4	ABG15232
10	6124.5	56.7	1129	8	ADQ66400
11	5026.5	46.5	1149	9	ADQ66400
12	5026.5	46.5	1149	9	ADQ66400
13	4506	41.7	897	8	ADQ66220
14	4488	41.6	897	8	ADQ66218
15	4494	41.6	897	8	ADQ66152
16	3694	34.2	734	7	ADQ66152
17	3694	34.2	734	7	ADQ66152
18	3102	28.7	624	9	AEEO2884
19	2585	23.9	515	8	AAW54099
20	2061	19.1	408	8	ADSD34906
21	2051.5	19.0	206	4	ABBS5242
22	2051.5	19.0	206	8	ADQ69606
23	1779	16.5	349	8	ADP24166

24	1778	16.5	349	4	AAB92729	AAB92729 Human pro
25	1603	14.8	325	7	ADQ69054	Adq69054 Novel pro
26	1603	14.8	325	9	ADU40573	Adu40573 Novel hum
27	1086	10.1	1798	6	ABJ26330	Abj26330 Aspergill
28	1085	10.0	1814	6	ABJ25730	Abj25730 Aspergill
29	1056.5	9.8	1818	5	ABP73779	Abp73779 Candida a
30	956	8.8	1769	6	ABR52974	AbR52974 Protein s
31	956	8.8	1769	7	ADK62658	AdK62658 Disease t
32	663	6.1	135	5	ADP41280	ADP41280 Human ova
33	434	4.0	437	8	ADX75081	Adx75081 Plant ful
34	379	3.5	77	4	AAW17368	Aaw17368 Peptide #
35	379	3.5	77	4	ABR36384	ABR36384 Peptide #
36	379	3.5	77	4	AAW29882	Aaw29882 Peptide #
37	379	3.5	77	4	ABB31185	Abb31185 Peptide #
38	379	3.5	77	4	AAW21738	Aaw21738 Protein #
39	379	3.5	77	4	AAW69542	Aaw69542 Human bon
40	379	3.5	77	4	AAW57146	Aaw57146 Human bra
41	379	3.5	77	4	ABG51215	Abg51215 Human liv
42	379	3.5	77	4	AAW05057	Aaw05057 Peptide #
43	379	3.5	77	5	ABG39167	Abg39167 Human pep
44	309.5	2.9	2228	7	ABR61599	ABR61599 Human gol
45	309.5	2.9	2230	6	ABU07445	Abu07445 Protein d

## ALIGNMENTS

RESULT 1  
ID AAB85029 standard; protein; 2144 AA.  
AC AAB85029;  
DT 06-AUG-2001 (first entry)  
XX  
XX Protein encoded by BAP28 cDNA consisting of exons 1 to 45.  
XX BAP28; prostate; tumour; cancer; diagnostic; genetic analysis.  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 1694 /label= Ser or Asn  
FT Misc-difference 1654 /label= Ala or Val  
FT Misc-difference 1967 /label= Asp or Asn  
FT Misc-difference 2017 /label= Gly or Glu  
XX  
XX WO200100669-A2.  
XX 04-JAN-2001.  
XX  
XX 23-JUN-2000; 2000WO-IB001183.  
XX  
XX 25-JUN-1999; 99US-0141323P.  
XX 18-JAN-2000; 2000US-0176880P.  
XX  
XX (GEST) GENSET.  
XX Barry C, Bouguetieret L, Chumakov I, Cohen-Akenine A;  
XX MPI; 2001-367032/38.  
XX N-PSDB; AAF83909, AAF83910.  
XX  
XX New BAP28 polynucleotides and polypeptides overexpressed in prostate  
XX cancer cells for diagnosing prostate tumors, e.g. by hybridization or  
XX polymerase chain reaction assays.  
XX Claim 14; Page 297-304; 349pp; English.

CC The invention is directed to BAP28 polypeptides, BAP28 polynucleotide  
CC sequences and regulatory region located at the 3' and 5' ends of the  
CC BAP28 coding region. The BAP28 polypeptides can be expressed by standard  
CC recombinant methodology. BAP28 polynucleotides and polypeptides have been  
CC found to be over expressed in prostate tumour cells, therefore levels of  
CC BAP28 expression and/or activity may be assayed (e.g. by polymerase chain  
CC reaction (PCR)) to diagnose patient suffering from or susceptible to  
CC prostate cancer. Antibodies specific for the BAP28 polypeptides are  
CC useful as diagnostic reagents. Biallelic markers of the BAP28 gene are  
CC encoded by a first CDNA sequence of the BAP28 gene consisting of the  
CC exons 1 to 45

XX Sequence 2144 AA:

Query Match 99.9%; Score 10799; DB 4; Length 2144;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MTSIAQQQLQRLALPQSDASLSRDEVASLFPPEKATIDRDTAFATGCTGLEELGIDP 60  
1 MTSIAQQQLQRLALPQSDASLSRDEVASLFPPEKATIDRDTAFATGCTGLEELGIDP 60  
61 SFEQFEAPLFSQAKTLERSVQTAKVANKQDENISLFIHSPYFLKPAQKCEMLIHR 120  
61 SFEQFEAPLFSQAKTLERSVQTAKVANKQDENISLFIHSPYFLKPAQKCEMLIHR 120  
121 FPHILYNODSLIACVLPHEHTRIFVRVIOQLKINNSKRMFMILPVKOSGVPILAKGTLIT 180  
121 FPHILYNODSLIACVLPHEHTRIFVRVIOQLKINNSKRMFMILPVKOSGVPILAKGTLIT 180  
121 FPHILYNODSLIACVLPHEHTRIFVRVIOQLKINNSKRMFMILPVKOSGVPILAKGTLIT 180  
181 HCYKDLGEMDFICSLVTKSVKFAEPGSSAQLRLALFAVASTIVSALVAEDVSDNIIA 240  
181 HCYKDLGEMDFICSLVTKSVKFAEPGSSAQLRLALFAVASTIVSALVAEDVSDNIIA 240  
181 HCYKDLGEMDFICSLVTKSVKFAEPGSSAQLRLALFAVASTIVSALVAEDVSDNIIA 240  
241 KLFPYIOGKKSIPDVRATATYMIICQISVKTMENTFVNSLAQIILKLIKISILKDG 300  
241 KLFPYIOGKKSIPDVRATATYMIICQISVKTMENTFVNSLAQIILKLIKISILKDG 300  
241 KLFPYIOGKKSIPDVRATATYMIICQISVKTMENTFVNSLAQIILKLIKISILKDG 300  
241 KLFPYIOGKKSIPDVRATATYMIICQISVKTMENTFVNSLAQIILKLIKISILKDG 300  
301 LSCILVILORQPSLSGKPPPHLCNVPLDTITILHGISETYDVSPLLRMYLPHLVSTIHH 360  
301 LSCILVILORQPSLSGKPPPHLCNVPLDTITILHGISETYDVSPLLRMYLPHLVSTIHH 360  
301 LSCILVILORQPSLSGKPPPHLCNVPLDTITILHGISETYDVSPLLRMYLPHLVSTIHH 360  
301 LSCILVILORQPSLSGKPPPHLCNVPLDTITILHGISETYDVSPLLRMYLPHLVSTIHH 360  
361 HTVGEETEGMDGQIYKQHLAAILTKISLKNLDDLHSLLEFYISVSSQEMDSNKYSL 420  
361 HTVGEETEGMDGQIYKQHLAAILTKISLKNLDDLHSLLEFYISVSSQEMDSNKYSL 420  
361 HTVGEETEGMDGQIYKQHLAAILTKISLKNLDDLHSLLEFYISVSSQEMDSNKYSL 420  
421 LNEQFLILIRLESKYPRTLDVLEBHLKETADLKQELPHQFVSLSTSGAKYQFLADSD 480  
421 LNEQFLILIRLESKYPRTLDVLEBHLKETADLKQELPHQFVSLSTSGAKYQFLADSD 480  
421 LNEQFLILIRLESKYPRTLDVLEBHLKETADLKQELPHQFVSLSTSGAKYQFLADSD 480  
481 TSLMLSLNHPALAPVRIILAMNHLKKIMKTSKEGVDSFIEKAVLARLDGDDIDVYLSAISA 540  
481 TSLMLSLNHPALAPVRIILAMNHLKKIMKTSKEGVDSFIEKAVLARLDGDDIDVYLSAISA 540  
481 TSLMLSLNHPALAPVRIILAMNHLKKIMKTSKEGVDSFIEKAVLARLDGDDIDVYLSAISA 540  
541 PEIYKEHPSSEVTISNLNLFQRAELSKNGEWEYLKADIIILKEETLSNDOLSNQVY 600  
541 PEIYKEHPSSEVTISNLNLFQRAELSKNGEWEYLKADIIILKEETLSNDOLSNQVY 600  
541 PEIYKEHPSSEVTISNLNLFQRAELSKNGEWEYLKADIIILKEETLSNDOLSNQVY 600  
541 PEIYKEHPSSEVTISNLNLFQRAELSKNGEWEYLKADIIILKEETLSNDOLSNQVY 600  
601 VCLLPFVAVINDDPESAEKIAIYLSKSGICSLPHLKGMEALENVIKSTKPGKLGIVA 660  
601 VCLLPFVAVINDDPESAEKIAIYLSKSGICSLPHLKGMEALENVIKSTKPGKLGIVA 660  
601 VCLLPFVAVINDDPESAEKIAIYLSKSGICSLPHLKGMEALENVIKSTKPGKLGIVA 660  
661 NOKMIELLADNINIGDPSMLKWEVDLISVGEESFNLKQVTFHVILISLVSCSSSLKE 720  
661 NOKMIELLADNINIGDPSMLKWEVDLISVGEESFNLKQVTFHVILISLVSCSSSLKE 720  
661 NOKMIELLADNINIGDPSMLKWEVDLISVGEESFNLKQVTFHVILISLVSCSSSLKE 720  
721 THRFPAIRVPSLLQKIKKLESVITTAVEIPEEMHILMDRGIVELMAHVEELNSQR 780  
721 THRFPAIRVPSLLQKIKKLESVITTAVEIPEEMHILMDRGIVELMAHVEELNSQR 780  
721 THRFPAIRVPSLLQKIKKLESVITTAVEIPEEMHILMDRGIVELMAHVEELNSQR 780  
781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMWLNCA 840  
781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMWLNCA 840  
781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMWLNCA 840

781 VAVDSVFLVSLKKFIYALKAPKSPKGDIMWNEQKEDSRDYLLHLILGFEMWLNCA 840  
841 DAHFVRLMKLFIKVLHEDVFOLEFKFCSVMTYTGSSLSNPLNGSVKTVLOTQALVYGCAM 900  
841 DAHFVRLMKLFIKVLHEDVFOLEFKFCSVMTYTGSSLSNPLNGSVKTVLOTQALVYGCAM 900  
841 DAHFVRLMKLFIKVLHEDVFOLEFKFCSVMTYTGSSLSNPLNGSVKTVLOTQALVYGCAM 900  
901 LSSQKTQCKQLASISSPVVTSLLINSGSVKEVRRRAIOTQALSGVASPFYIIDLHI 960  
901 LSSQKTQCKQLASISSPVVTSLLINSGSVKEVRRRAIOTQALSGVASPFYIIDLHI 960  
901 LSSQKTQCKQLASISSPVVTSLLINSGSVKEVRRRAIOTQALSGVASPFYIIDLHI 960  
961 SKAEITSDAAVYIOTDALTPEELOREKUKSHOKSETLKNILSCVYSCPSYAKDLMK 1020  
961 SKAEITSDAAVYIOTDALTPEELOREKUKSHOKSETLKNILSCVYSCPSYAKDLMK 1020  
961 SKAEITSDAAVYIOTDALTPEELOREKUKSHOKSETLKNILSCVYSCPSYAKDLMK 1020  
1021 VLOGVNGEMVLSQLPMAEOLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080  
1021 VLOGVNGEMVLSQLPMAEOLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080  
1021 VLOGVNGEMVLSQLPMAEOLLEKIQKEPTAVLKDEAMVHLTLGKXNEFSVSLNEDPK 1080  
1081 SLDFIKAVHTTKELVAGMPTIQTALAKITKPPFAAISPEKYOQKILRMLPDLVNCN 1140  
1081 SLDFIKAVHTTKELVAGMPTIQTALAKITKPPFAAISPEKYOQKILRMLPDLVNCN 1140  
1081 SLDFIKAVHTTKELVAGMPTIQTALAKITKPPFAAISPEKYOQKILRMLPDLVNCN 1140  
1141 SHCAQTVSSVFKGISVNAEQVRLEBPDKAKPLGTVOQRROKQKQKSDLESVQEVG 1200  
1141 SHCAQTVSSVFKGISVNAEQVRLEBPDKAKPLGTVOQRROKQKQKSDLESVQEVG 1200  
1141 SHCAQTVSSVFKGISVNAEQVRLEBPDKAKPLGTVOQRROKQKQKSDLESVQEVG 1200  
1201 GSYWQVRLILLELLOHKKRLSPQIIVPTLNLNLSRCLBPLPEQGNMEYTKOLILSCLL 1260  
1201 GSYWQVRLILLELLOHKKRLSPQIIVPTLNLNLSRCLBPLPEQGNMEYTKOLILSCLL 1260  
1201 GSYWQVRLILLELLOHKKRLSPQIIVPTLNLNLSRCLBPLPEQGNMEYTKOLILSCLL 1260  
1261 NIOQKLSPDGDKIPKOLIDBEKENVELIYOCIRLSEPOTHHALLLGVAGIFPPKVL 1320  
1261 NIOQKLSPDGDKIPKOLIDBEKENVELIYOCIRLSEPOTHHALLLGVAGIFPPKVL 1320  
1261 NIOQKLSPDGDKIPKOLIDBEKENVELIYOCIRLSEPOTHHALLLGVAGIFPPKVL 1320  
1321 HNIMSIPTFMGAVNMLDDTYSFOVINKTVKQVIPALIQSDSGDSEVSNNVEIIVKII 1380  
1321 HNIMSIPTFMGAVNMLDDTYSFOVINKTVKQVIPALIQSDSGDSEVSNNVEIIVKII 1380  
1321 HNIMSIPTFMGAVNMLDDTYSFOVINKTVKQVIPALIQSDSGDSEVSNNVEIIVKII 1380  
1381 SVEFDALPHVPEHRRRLPILVOLVDTGAKEFLWILILLFEOYVTVKTVLAAVGEKDAIL 1440  
1381 SVEFDALPHVPEHRRRLPILVOLVDTGAKEFLWILILLFEOYVTVKTVLAAVGEKDAIL 1440  
1381 SVEFDALPHVPEHRRRLPILVOLVDTGAKEFLWILILLFEOYVTVKTVLAAVGEKDAIL 1440  
1441 EADTEPFWSVCCSFVOHQOSLMLNLIQVLLKLPKEKEETIPRAVSPKSESQEMLOVF 1500  
1441 EADTEPFWSVCCSFVOHQOSLMLNLIQVLLKLPKEKEETIPRAVSPKSESQEMLOVF 1500  
1441 EADTEPFWSVCCSFVOHQOSLMLNLIQVLLKLPKEKEETIPRAVSPKSESQEMLOVF 1500  
1501 NVEHTSKQLRHFKFLSVSFMSQLSSNNFLKRVESGGEIILKGLBERLLETVLGYISA 1560  
1501 NVEHTSKQLRHFKFLSVSFMSQLSSNNFLKRVESGGEIILKGLBERLLETVLGYISA 1560  
1501 NVEHTSKQLRHFKFLSVSFMSQLSSNNFLKRVESGGEIILKGLBERLLETVLGYISA 1560  
1561 VAOSMERADKLTVKWTRALLSKQYDLDBKVNALLPTEPTIPYIRGLVGNPLASVRKAL 1620  
1561 VAOSMERADKLTVKWTRALLSKQYDLDBKVNALLPTEPTIPYIRGLVGNPLASVRKAL 1620  
1561 VAOSMERADKLTVKWTRALLSKQYDLDBKVNALLPTEPTIPYIRGLVGNPLASVRKAL 1620  
1621 DLNNKLOQNIWKKITVTRFLKLVDPDLAIYORKKKEGEEOAIRQATLYTLKLCKN 1680  
1621 DLNNKLOQNIWKKITVTRFLKLVDPDLAIYORKKKEGEEOAIRQATLYTLKLCKN 1680  
1621 DLNNKLOQNIWKKITVTRFLKLVDPDLAIYORKKKEGEEOAIRQATLYTLKLCKN 1680  
1681 FGAENDDPFPVVLXTAVKLJAPERKEKNVLSGALLCIAEVTSTLEALAIPLPBLMPSL 1740  
1681 FGAENDDPFPVVLXTAVKLJAPERKEKNVLSGALLCIAEVTSTLEALAIPLPBLMPSL 1740  
1681 FGAENDDPFPVVLXTAVKLJAPERKEKNVLSGALLCIAEVTSTLEALAIPLPBLMPSL 1740  
1741 LTTMKNTSELVSSEVYLLSLAALQKVEVLPPIISYLEGILISQVHLHEKITSKMSAS 1800  
1741 LTTMKNTSELVSSEVYLLSLAALQKVEVLPPIISYLEGILISQVHLHEKITSKMSAS 1800  
1741 LTTMKNTSELVSSEVYLLSLAALQKVEVLPPIISYLEGILISQVHLHEKITSKMSAS 1800  
1801 QANIRLSTLKKTLATTLAPRVLPAIKTYKQIEKNWKKNMGPFMSILOHIGMKKKEEL 1860  
1801 QANIRLSTLKKTLATTLAPRVLPAIKTYKQIEKNWKKNMGPFMSILOHIGMKKKEEL 1860  
1801 QANIRLSTLKKTLATTLAPRVLPAIKTYKQIEKNWKKNMGPFMSILOHIGMKKKEEL 1860  
1861 TSHQSOLTAFLBALDFRAQHSNDEEVGKTENCIIDCLVAVVYKLSVTFRRPLPKLF 1920  
1861 TSHQSOLTAFLBALDFRAQHSNDEEVGKTENCIIDCLVAVVYKLSVTFRRPLPKLF 1920  
1861 TSHQSOLTAFLBALDFRAQHSNDEEVGKTENCIIDCLVAVVYKLSVTFRRPLPKLF 1920



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:42:00 ; Search time 51.7106 Seconds  
(without alignments)  
3989.295 Million cell updates/sec

Title: US-09-603-665-5

Perfect score: 10803  
Sequence: 1 MTSLAQQLRLALPQSDASL.....CCKTIQLETVLGEPLQSYF 2144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990.5	9.2	1649	2	T39938 hypothetical prote
2	956	8.8	1769	2	S53378 probable membrane
3	857	7.9	1650	2	T27864 hypothetical prote
4	317.5	2.9	2670	2	T37919 GCN1 homolog - fib
5	279	2.6	2672	2	A48126 translation activa
6	272.5	2.5	2954	2	T14156 kinesin-related pr
7	268.5	2.5	2663	1	S28261 centromere protein
8	261.5	2.4	1727	2	A90551 conserved hypotet
9	260	2.4	1727	2	T50073 myosin-like coiled
10	257	2.4	4385	2	T29042 hypothetical prote
11	256.5	2.4	2937	2	AB2894 hypothetical prote
12	255.5	2.4	1837	2	T41023 probable nuclear p
13	254.5	2.4	3433	1	S28381 utrophin - human
14	252	2.3	2895	2	H85362 hypothetical prote
15	249.5	2.3	3655	2	T38084 TRAP-like protein
16	249	2.3	3187	2	JCS837 364K Golgi complex
17	249	2.3	2401	2	T28676 rhoptry protein -
18	247	2.3	2493	2	T40540 hypothetical prote
19	245	2.3	2059	2	T41933 large tegument pro
20	244.5	2.3	2514	2	T37320 ataxia telangiecta
21	244	2.3	2110	2	H96803 unknown protein T5
22	243	2.2	2269	2	T28677 rhoptry protein -
23	242	2.2	1964	2	A59282 nonmuscle myosin I
24	238	2.2	1957	2	T38077 hypothetical coile
25	236.5	2.2	1830	2	B82809 conserved hypotet
26	235.5	2.2	2471	2	T42971 large tegument pro
27	235	2.2	1919	2	T40032 hypothetical prote
28	234	2.2	3795	2	T00831 hypothetical prote
29	232.5	2.2	2166	2	G70163 hypothetical prote

30	232.5	2.2	2470	2	S57085 1-phosphatidylinos
31	232.5	2.2	2712	2	T05113 hypothetical prote
32	231	2.1	2253	2	T30336 nuclear/mitotic ap
33	230	2.1	2335	2	T40186 probable phosphati
34	229.5	2.1	2104	2	T38774 myosin-3 heavy cha
35	228	2.1	4131	2	T21085 hypothetical prote
36	227	2.1	1922	2	T00637 hypothetical prote
37	227	2.1	1939	2	T18372 repeat organellar
38	226	2.1	1447	2	F82909 hypothetical prote
39	225.5	2.1	1679	2	S48385 giantin - human
40	225.5	2.1	3225	2	T52300 giantin - human
41	225.5	2.1	3259	1	A56539 giantin - human
42	224.5	2.1	1819	2	A71928 cag island protein
43	224	2.1	1959	1	A33977 myosin heavy chain
44	224	2.1	2748	2	S57976 nuclear migration
45	224	2.1	3738	2	T05501 hypothetical prote

ALIGNMENTS

RESULT 1	
T39938	hypothetical protein SPBC23B.04C - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe	
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004	
C/Accession: T39938	
R/Wood, V.; Rajandream, M.A.; Barrell, B.G.; Pohl, T.	
A.Submitted to the EMBL Data Library, May 1998	
A.Reference number: Z21892	
A/Accession: T39938	
A/Status: preliminary; translated from GB/EMBL/DBJ	
A/Molecule type: DNA	
A/Residues: 1-1649 <WOO>	
A/Cross-references: UNIPROT:O60179; UNIPARC:UP1000013AF93; EMBL:AL023287; PIDN:CAA18872.1	
C.Experimental source: strain 972h-; cosmid c23B6	
C.Genetics:	
A/Gene: SPDB:SPBC23B.04C	
A/Map position: 2	
Query Match	9.2%; Score 990.5; DB 2; Length 1649;
Best Local Similarity	21.1%; Pred. No. 2e-40;
Matches 469; Conservative 379; Mismatches 718; Indels 655; Gaps 78;	
QY	2 TSLAQQLRLALPQSDASLSRDEV---ASLLFDPEKATITRDTRFAIGCTGLEBLGI 58
DB	3 SSIQKQLKNI---QSN-NVLKINKIRAPSLIYDPKVAADMLEEIVYVAVGFEHLAVH 58
QY	59 DSFFQFEAPLFSQAKTLERSVQTKAVNKKOLDENISLFLHSPYELKPKQKLEWLI 118
DB	59 EPRLLYFEKTLIGEOSVOQDRLVLTNTEKIDLECVOQLRLLAPEFTTKNAKLVLEMLI 118
QY	119 HFFHILYNQDSLIVACVLYPHETRIFFRVYQLKINKNSKGRFWLLPVKQSGVPLAKGTL 178
DB	119 RRFSIHEVSDFFILSLFEPHDHPFARILGCSK-PKSRPLFLFLENAIKMP-VTISRADI 176
QY	179 ITHCYKDLGFMDPFCISLVTKSVKVAEYFGSSAQRLVLAFASTIVSLVAEDVSDN- 237
DB	177 VHALSRDKKEFFMFAQFOVNTAESHNMP-----ELARFWAGTMEVLVAMHSSNEDP 229
QY	238 --IIAKLPFYIQKGLK--SLPDYRATATMIIQISVKTMENTFVNSLASQIKTLTKI 293
DB	230 NVLLDRFFLRVAVSVYSI-DFOIAGMLSSIIAASIPSPSIIPPLVSAITDLSF- 287
QY	294 PELINDGSLCLIVL-----QKQPSISLQ----- 317
DB	288 -DNIKPALICVGHLLQFCSSFEFDHQLEKTESFGASSLIELSQHRIDFEFVSYVSL 346
QY	318 -----KKPPPHCANPBDLITILAGISSETVDSPFLRYMPLHVSIHHVTEGTEGMDG 372
DB	347 IKSRRQKDKKRILSLD--TSISQIRVTHEQAKFLSLVP-----VNDDFAL-- 392
QY	373 QYKSHLEAILTKISLKNLHLASLLFEFYISYSSQREMSNKVSLNEQFLPLIRLL 432



Db	393	QSYRRLDVSVOPEPKSGKLDNLNTL	----	ODKKXSSTFKKPREVL	----	436
Oy	433	ESKYPRTLDVLEBHLXELADLKQOELFHQFVSLSTGCGKYQFLADSDTSLMLNPLA	492			
Db	437	-----LKKISEIDQTSPEOGLAADA	----	ADDSVFLSL	----	470
Oy	493	PVRLANMHLKKIMKTSKEGVDESFIKEAVLARIADNDIVVLSAISAFELIKENHSSEV	552			
Db	471	-----LSKEG-DKIPFLLCIA	----	NGSERI	----	492
Oy	553	TISMLNLFORAELSKNGEWEYVLKIADILIKEIILISENDOLSNQVCLLPEV	----	-IN	610	
Db	493	IILSIEL	-----	RKITEKNDVYOII	----	LPVLVLSLO 524
Oy	611	NDPTESEMKIAIYLSKSGICSLHPLRGWEALENVIKSTPKGLIGVANQMIELAD	670			
Db	525	SKOTEVRSRALNLI	----	-----	FLEIRNE	547
Oy	671	NINIGDSMLKMWEDLSVGEESFNKKQVTPHVLISLVSCSSKLKEHPFAIRVF	730			
Db	548	NL	-----	EFSTIYMDNDN	----	562
Oy	731	SLQOKIKLESVTAVAEIPSEWHIELMDRGIPVEIMAHYEELNSTORVAVESVFLV	790			
Db	563	-----KNLRWLSPEV	-----	KYCSDDLDRSSEIGL	----	DGYLTF 596
Oy	791	FSLKFFIYALAPKPSFPKGDIMWNPBQLEKDSRDYHLHIGLFEMLMNGADAHFRVLK	850			
Db	597	SYIPERLETEKPKPKXSK	-----	-----	ELAVT	619
Oy	851	LFIKHLEDDVQLFKFCSVLWTYSSLSNPINCQVKTVLQTOALYGCAMLSOKTOCKH	910			
Db	620	SFLSSH	-----	-----	AACSKLSN	633
Oy	911	QLASISFVVTSLNLNGSPVKEVRRRAIOCLQALSGVASFYLIIDHLISKAETISDA	970			
Db	634	-----	-----	RVLLLEILTRVHG	----	KVED 651
Oy	971	AVVIQDALTPEELOREBKKLSHOKLSEBTLKNLISCVYSCPSYIAKDLMLKLOCVNGEMV	1030			
Db	652	-----AKXOILPLRLEQSEFNS	-----	EKFKTVSKREBALVNCNHRIS	----	FTLSLSTLSNIV 704
Oy	1031	LSOLLPMABOLLEKIOKEPVALKDQEAUHLTLGKYNESFVSLINDEPKSLDIFIAVH	1090			
Db	705	LSQAI-CRRIIVE-IO	-----	HL	----	KDPRLE-FVYAVI 733
Oy	1091	TTKEIYAGMPITQITALEKITKRPFAAISDEKVQCKLRLMLFDLLVNCNKHCAQOTVSV	1150			
Db	734	SODE	-----	OPHYVDVLDISIKIP	----	DTV 755
Oy	1151	FKGISVNAEQVRIELEPPDKAKPLGTVOOKRROKQOKKSODELVGESSWQWVTLI	1210			
Db	756	FK	-----	KLGSVALVEKROPAIAKRRDRS	----	HIFGDVORLTRI 794
Oy	1211	LELIQHKKLRSPOILVPTLNLRSCLPELPQEOGMEYTKOILISCLNI	-----	CQK	1265	
Db	795	LELEETKNAASYPKLASP-LFEVLSNV	-----	ALKEDIVSSNYLLOLILGLYEMIGASPITE	852	
Oy	1265	LSPOGKIPKIDLEKENEVELIYOCIRLSLEMPQTHHALLLGTVAGIIPDKYLIHNIMS	1325			
Db	853	LSP	-----	SIRIDTIVOCISISTNNPOJONKALLVSLANAAAEVALHGMV	899	
Oy	1326	IFTEMGANVMRLDITYSFQVINKTYKMIIPALIQ	-----	SUSGSIEXSRVBEIIVKIISVF	1383	
Db	900	IFTEGGSFVLSRDPDAFSLHVI	-----	EQTAKTVISALIRIGKDPDSL	----	LVSCEF 947
Oy	1384	VDALPHVEHRLPIVLQVLDVDTLGAEKFEIMILLILFEQVYTKVILAAAGEKALIEAD	1443			
Db	948	VNAFPHIQHRLRLXRLVLOTIGSNRFLSVLI	-----	QPAEKMLLA	----	KSTNVVAI 998
Oy	1444	TEFWSVCCFESVOHQIOSLNMNILOYLLKLPBEK	-----	EETIPKAVSFNKSQOEMLOVEN	1501	

```

Db 999 HFCFLTLVQSSVADPRISINQCSRFCCLKSLBEQNSDNGKAVSLIKUDELP-----MD 1053
Qy 1502 VETHTSKOLRHFKFLSVFSMGOLLSNNP---LKKVYESGCEILKGLBERLLETVLGYI 1558
Db 1054 VDLATLGLSLR-VKVEL--ISLVSRAKNFAPFLAKIMENSVDSPVE-IOAGFJESIKLLI 1109
Qy 1559 SAVAS---MERNDKLTVMKFWBALSKAYDLDDKYNALLPETFPFIPVRGLGNPLPSV 1615
Db 1110 TLSQOOSNEME-----LGHVYVALRSVITHLLPHELFPCTVGLKLLHDERALL 1155
Qy 1616 RRKALDILNNKLQONI SWKKTIVTRFLKLVPLDLAI VOKKKKEGEEQAINROTALYTLK 1675
Db 1156 RRKALSTI VQOQVQOQ-----SKVSALTALIPD---VTYINISVSDDE---TTLADNCLA 1204
Qy 1676 LLCNKGEGENDDPFPVLXTAVKLTAPB---RKEEKNVGSGALLCJAEVYSTLEALAIPO 1732
Db 1205 VMAKFS-SPELF---SPLEVVSGPGLKNSARDVQVSAIVCTIVLTNTLTAARILPY 1259
Qy 1733 LPSLMPSLITTMKNTSELVSEVYLLSALALQKVEVLPHPFISPLYEGILSOVIHLEKI 1792
Db 1260 LADIYVNSTI LIDARKDPEGDLLELACFSNMIDFPKVPPEFSSSYVEPTIKCALASDR- 1318
Qy 1793 TSEMSASQAINRLTSLKKTATTAPRYLPAIKTYQJIEKNMKNHGPFPMSIOEHI 1852
Db 1319 -----AEHDDIGEGLPETTINFIPTRLKMSIPRAWEBCARLGSTALRLLELEBAL. 1372
Qy 1853 GXMKEELTSHOSQUTAFLEALDFRAQHSENDLEBVGKTENCIIDCLVAMVVKLSVTF 1912
Db 1373 QNSRSHSIGTYVKSIFKEFFLDSFDSR--RSLFPAEDVDVNETQAVNVFLKFWKLSPTTF 1430
Qy 1913 RPLFPKLDMAKTEBAPKD-----RLTLFYVALDCAIAKLGFLPLFGHLVAKFPADTL 1966
Db 1431 RPLFHLHSMALBEDYETDPSGIVSRQTFYFPLTIFLDTLKSIVNNYAAVYD---DTI 1487
Qy 1967 XQVNI SKTDEAFDSENDPEKCCLLLOFLINCLYKILFDPDGHF-ISKERAXXALMPLVD 2025
Db 1468 ELSSKOT-----NSE-----VRHLVNSSLSVAFENDTEEFMVVPARFGKISPVLIIE 1534
Qy 2026 QLENRLGEBEKEQENVTHKLI PCIAQF-SVAMADSLIMPLVNOILLKTPDSSPKVFAA 2084
Db 1535 QIYVAPLIDDKX-----LVRAIVELASVASASSSD- FFSMNTQILOLYLRSSINARLLA 1586
Qy 2085 LITVALAEKLEKENYIVLLPESIPFLAEI MEDECEVEHQOCTIOOLETVLG--EPLOS 2142
Db 1587 IQIQOQLVGRGEMWISTLPQSVFPFIAELMEDDDQVETATAELVRIIDRLGENESLQD 1646
Qy 2143 Y 2143
Db 1647 Y 1647

```

RESULT 2  
S53378  
probable membrane protein YJL109c - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein J0808  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 05-May-1995 #sequence revision 01-Sep-1995 #text\_change 09-Jul-2004  
A:Accession: S53378; S56887; S57359  
R:Raumussen, S.W.  
submitted to the EMBL Data Library, February 1995  
A:Description: A 37.5 kb region of yeast chromosome X includes the SMI1, MEF2, GSH1 and C  
A:Reference number: S53376  
A:Accession: S53378  
A:Molecule type: DNA  
A:Residues: 1-1769 >RAS>  
A:Cross-references: UNIPROT: P42945; UNIPARC:UPI00000052F40; EMBL: X85021; NID: G728698; PID:  
R:Raumussen, S.W.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56876  
A:Accession: S56887  
A:Molecule type: DNA  
A:Residues: 1-1769 >RAM>  
A:Cross-references: UNIPARC:UPI00000052F40; EMBL: Z49384; NID: G10008292; PID: G10008293; MIPS

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:41:34 ; Search time 219.344 Seconds

(without alignments)

9041.668 Million cell updates/sec

Title: US-09-603-665-5

Sequence: 1 MSLAQQQLRLALPQSDASL.....CCKTQQLEWVGEPIQSYF 2144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : UniProt 7.2.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10791	99.9	2144	1 HEATR_HUMAN	09h83 homo sapien
2	10342.5	95.7	2063	2 O5T3O7_HUMAN	05t3q7 homo sapien
3	5807.5	53.8	2159	2 O7S148_BRARE	07s148 brachydanto
4	5562	51.5	1106	2 O6P197_HUMAN	06p197 homo sapien
5	4944	45.8	2288	2 O4T723_TETNG	04t723 tetradon n
6	4891	45.3	1180	2 O3UNW7_MOUSE	03unw7 mus musculu
7	4729	43.8	958	1 HEATR_MACRA	09gm44 macaca fasc
8	4516	41.8	1090	2 O3TNEI_MOUSE	03tnei mus musculu
9	4494	41.6	897	2 O8N1L7_HUMAN	08n1l7 homo sapien
10	4072	37.7	984	2 O3TKMO_MOUSE	03tkm0 mus musculu
11	3935.5	36.4	1336	2 O7T152_BRARE	07t152 brachydanto
12	3669.5	34.0	1378	2 O7T153_BRARE	07t153 brachydanto
13	3315.5	30.7	743	2 O3V1X6_MOUSE	03v1x6 mus musculu
14	3099	28.7	733	2 O8BLJ4_MOUSE	08blj4 mus musculu
15	3085	28.6	733	2 O3F1910_MOUSE	03f1910 mus musculu
16	2156.5	20.0	2104	2 O7PMD6_ANOGA	07pmd6 anopheles g
17	2051.5	19.0	2096	1 HEATR_DROME	09vm75 drosophila
18	2051.5	19.0	2096	2 O6AMS0_DROME	06ams0 drosophila
19	1926	17.8	408	2 O8CCT5_MOUSE	08cct5 mus musculu
20	1779	16.5	349	2 O9E6S5_HUMAN	09e6s5 homo sapien
21	1662	15.6	349	2 O8VCK1_MOUSE	08vck1 mus musculu
22	1462.5	13.5	1590	2 O8T9E7_DROME	08t9e7 drosophila
23	1415	13.1	2237	2 O54ML4_DICDI	054ml4 dictyosteli
24	1252	11.6	2122	2 O5CAF8_ORYSA	05caf8 oryza sativ
25	1194.5	11.1	1830	1 HEATR_ARATH	09kc2 arabidopsis
26	1142	10.6	2021	2 O5KBO2_CRYNE	05kbo2 cryptococcu
27	1134	10.5	2021	2 O5SNB7_CRYNE	05snb7 cryptococcu
28	1127.5	10.4	1801	2 O5B1X5_EMBNI	05b1x5 aspergillus
29	1127	10.4	1857	2 O6BXO6_DEBHA	06bxo6 debaryomyce
30	1104	10.2	1802	2 O2UIC6_ASPOR	02uic6 aspergillus
31	1088	10.1	1798	2 O4WLI9_ASPFU	04wli9 aspergillus

32	1056.5	9.8	1818	2 O592X6_CANAL	0592x6 candida alb
33	1049.5	9.7	2251	2 O4P937_USTWA	04p937 ustilago ma
34	1030.5	9.5	1788	2 O7R2M8_NEUCR	07r2m8 neurospora
35	1006.5	9.3	1770	2 O6FRJ3_CANCA	06frj3 candida gla
36	1003.5	9.3	1774	2 O6CJ57_KLUJA	06cj57 kluyveromyc
37	990.5	9.2	1649	1 YG44_SCHPO	060179 schizosacch
38	965	8.9	1806	2 O41696_GIBZE	041696 gibberella
39	956	8.8	1769	1 UTP10_YEAST	P42945 saccharomyc
40	950.5	8.8	1774	2 O754J8_ASHGO	0754j8 ashyia goss
41	903	8.4	1635	2 O6C457_YARLI	06c457 yarrowia li
42	870.5	8.1	1660	2 O61Y68_CABER	061y68 caenorhabdi
43	857	7.9	1650	1 HEATR_CABEI	023495 caenorhabdi
44	852	7.9	168	2 O2KIF9_BOVIN	02kif9 bos taurus
45	773	7.2	2034	2 O519S9_ENTHI	0519s9 entamoeba h

## ALIGNMENTS

RESULT 1  
HEATR\_HUMAN STANDARD; PRT; 2144 AA.  
ID O5H583; O5T3O8; O9NM23;  
AC 09H583; O5T3O8; O9NM23;  
DT 01-JUN-2001. Integrated into UniProtKB/Swiss-Prot.  
DT 19-JUL-2005. sequence version 3.  
DT 07-MAR-2006. entry version 33.  
DE HEAT repeat-containing protein 1 (Protein BAP28).  
GN Name=HEATK1; Synonyms=BAP28;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE, AND VARIANTS SER-1694; ALA-1854; ASP-1967 AND  
RP GLY-2017.  
RA Bouguetel et al., Chumakov I., Berty C., Cohen-Akenine A.;  
RT "A novel BAP28 gene and protein."  
RL Patent number WO0100669, 04-JAN-2001.  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RG Human chromosome 1 international sequencing consortium;  
RL Submitted (May-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 1777-2144.  
RX PubMed:14702039; DOI:10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine N., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsura N., Sato K., Tanigami M., Yamazaki M.,  
RA Niimiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tani H., Kimura M., Watanabe M., Hirano S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanohori K., Takahashi-Fuji A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togawa S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Mushino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikemura Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45 (2004).  
RN (4)  
RP IDENTIFICATION BY MASS SPECTROMETRY AND SUBCELLULAR LOCATION.  
RX MEDLINE=22317277; PubMed=12429849; DOI=10.1091/mbc.E02-05-0271.  
RA Greco A., Conte Y., Deon C., Calle A., Kindbelter K., Sanchez J.-C.,  
RA Sgreto A., Hochstrasser D.F., Diaz J.-J.;  
RL "Functional proteomic analysis of human nucleolus.";  
RL Mol. Biol. Cell 13:4100-4109 (2002).  
CC -1- INTERACTION:  
CC 06W67: XEN1; IncAct=EBI-373098; EBI-372406;  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -1- SIMILARITY: Belongs to the HEATR1 family.  
CC -1- SIMILARITY: Contains 1 HEAT repeat.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonDerivs License  
CC -----  
CC EMBL: AX067150; CAC26776.1; -; Unassigned DNA.  
CC EMBL: AL359921; CA133775.1; -; Genomic DNA.  
CC EMBL: AK001221; BAA91564.1; ALT\_INIT; tRNA.  
CC INACC: Q9H583;  
CC SWISS-2DPAGE: Q9H583; HUMAN.  
CC Ensembl: ENSG00000119285; Homo sapiens.  
DR HGNC: HGNC:25517; HEATR1.  
DR GO: GO:0005515; F:protein binding; IPI.  
DR InterPro: IPR012954; BP28 C.  
DR InterPro: IPR00357; HEAT.  
DR Pfam: PF08146; BP28CT; 1.  
DR Pfam: PF02988; HEAT; 4.  
DR PROSITE: PS00077; HEAT\_REPEAT; FALSE\_NEG.  
KW Nuclear protein; Polymorphism.  
FT CHAIN 1 2144 HEAT repeat-containing protein 1.  
FT /FTID=PRO\_0000186201.  
FT REPEAT 2106 2142 HEAT.  
FT VARIANT 1694 1694 N -> S.  
FT /FTID=VAR\_010939.  
FT VARIANT 1854 1854 V -> A.  
FT /FTID=VAR\_010940.  
FT VARIANT 1967 1967 N -> D.  
FT /FTID=VAR\_010941.  
FT VARIANT 2017 2017 E -> G.  
FT /FTID=VAR\_010942.  
FT CONFLICT 348 348 H -> R (in Ref. 1).  
FT CONFLICT 607 607 M -> V (in Ref. 1).  
SQ SEQUENCE 2144 AA; 242370 MW; 7590EBA17A5FF39D CRC64;  
Query Match 99.9%; Score 10791; DB 1; Length 2144;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2138; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 MTSLAQQLORLALQSPDASLSRDEVASLSLFDPKKAATIDRDTAFAGCTGLEELGIDP 60  
DB 1 MTSLAQQLORLALQSPDASLSRDEVASLSLFDPKKAATIDRDTAFAGCTGLEELGIDP 60  
QY 61 SFEQFEAPLSQALKTLERSVOTKAVNKKDENTISLFLHLSPYFLLPKPAQCLEMLIHR 120  
DB 61 SFEQFEAPLSQALKTLERSVOTKAVNKKDENTISLFLHLSPYFLLPKPAQCLEMLIHR 120  
QY 121 FHHHLNODSLIAACVLPYHETRIFFVRVQLTKINNSKRMFWLLPVKOSGVPLAKGTLIT 180  
DB 121 FHHHLNODSLIAACVLPYHETRIFFVRVQLTKINNSKRMFWLLPVKOSGVPLAKGTLIT 180  
QY 121 FHHHLNODSLIAACVLPYHETRIFFVRVQLTKINNSKRMFWLLPVKOSGVPLAKGTLIT 180  
DB 121 FHHHLNODSLIAACVLPYHETRIFFVRVQLTKINNSKRMFWLLPVKOSGVPLAKGTLIT 180  
QY 181 HCYKDLGFMDFCISLVTKSVKFAEPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240  
DB 181 HCYKDLGFMDFCISLVTKSVKFAEPGSSAQLRVLLAFVASTIVSALVAEDVSDNIIA 240  
QY 241 KLFPIYIQGLKSLSPDRAATYMIICQISVAYTMENTVNSLASQIIKTLTKISLXDG 300  
DB 241 KLFPIYIQGLKSLSPDRAATYMIICQISVAYTMENTVNSLASQIIKTLTKISLXDG 300  
QY 241 KLFPIYIQGLKSLSPDRAATYMIICQISVAYTMENTVNSLASQIIKTLTKISLXDG 300  
DB 241 KLFPIYIQGLKSLSPDRAATYMIICQISVAYTMENTVNSLASQIIKTLTKISLXDG 300

QY 301 LSCLIYLLORQKPEISGKKPFPPLCNVPPDLITLHGISETYDVSPLLRYMLPHLVYSIIH 360  
DB 301 LSCLIYLLORQKPEISGKKPFPPLCNVPPDLITLHGISETYDVSPLLRYMLPHLVYSIIH 360  
QY 361 HVTGEETEGMDGQIYKRHEALITKISLKNNDLHLASLFFEEIYISYSSOEEDSKNVS 420  
DB 361 HVTGEETEGMDGQIYKRHEALITKISLKNNDLHLASLFFEEIYISYSSOEEDSKNVS 420  
QY 421 LNEQPLPLRLSESKYPRTLDDVLEBHLKEIADLKKQELPHQFVSLSTSGKYQPLASD 480  
DB 421 LNEQPLPLRLSESKYPRTLDDVLEBHLKEIADLKKQELPHQFVSLSTSGKYQPLASD 480  
QY 481 TSLMLSLNPLAVRLLANHLKIKMTSKEGVDESPFKAVLARIGDNDIVVLSAISA 540  
DB 481 TSLMLSLNPLAVRLLANHLKIKMTSKEGVDESPFKAVLARIGDNDIVVLSAISA 540  
QY 541 FEIFKHFSESEVTISULNLFORAELSKNGEYEVKIAADILKEEILSENDQLSNOV 600  
DB 541 FEIFKHFSESEVTISULNLFORAELSKNGEYEVKIAADILKEEILSENDQLSNOV 600  
QY 601 VCLLPVAVINDDTESAEMKIAIYLSKSGICSHPLLRGMEALENVIKSTKPKGLIGVA 660  
DB 601 VCLLPVAVINDDTESAEMKIAIYLSKSGICSHPLLRGMEALENVIKSTKPKGLIGVA 660  
QY 661 NQKMIELADNINLGDSSWLKXVEDLISGEESEFNKKQVFFHYILSVLSCSSLKE 720  
DB 661 NQKMIELADNINLGDSSWLKXVEDLISGEESEFNKKQVFFHYILSVLSCSSLKE 720  
QY 721 THPPFAIRVSLQKKIKKLESYITAVEIPSEWHIELMDRGIPVELMAHYEELNSTOR 780  
DB 721 THPPFAIRVSLQKKIKKLESYITAVEIPSEWHIELMDRGIPVELMAHYEELNSTOR 780  
QY 781 VAVEDSVYFLVFSIKKTIYALKAKSPKPGDIWNPOLKEDSDYHLHLIGLEMMNLGA 840  
DB 781 VAVEDSVYFLVFSIKKTIYALKAKSPKPGDIWNPOLKEDSDYHLHLIGLEMMNLGA 840  
QY 841 DAVHPRVIMKLFKXHLBEDVFOLFKCSVMTGSSLSNPINCSVKTVLOTQALYVGCAM 900  
DB 841 DAVHPRVIMKLFKXHLBEDVFOLFKCSVMTGSSLSNPINCSVKTVLOTQALYVGCAM 900  
QY 901 LSSQKQCKQLASISSPVVTSLLINLGSFVKEVRRAIQCIALSGVASPFYLIIDHLI 960  
DB 901 LSSQKQCKQLASISSPVVTSLLINLGSFVKEVRRAIQCIALSGVASPFYLIIDHLI 960  
QY 961 SKABEITSDAAYIOTLATLFELOREKLLKSHQKSETLKNLLSCVYSCPSTIADLMK 1020  
DB 961 SKABEITSDAAYIOTLATLFELOREKLLKSHQKSETLKNLLSCVYSCPSTIADLMK 1020  
QY 1021 VLOGVNGEAWLSQLLPMAEQLEKIQKEPTAVLKDEAMVHLTLGKYNBSVSLNEDPK 1080  
DB 1021 VLOGVNGEAWLSQLLPMAEQLEKIQKEPTAVLKDEAMVHLTLGKYNBSVSLNEDPK 1080  
QY 1081 SLDIFIKAVHTTKELYAGMPTIQITALEKIKTPPPAISDEKYQOKLLMLPDLVNNCKN 1140  
DB 1081 SLDIFIKAVHTTKELYAGMPTIQITALEKIKTPPPAISDEKYQOKLLMLPDLVNNCKN 1140  
QY 1141 SHQAQVTSYFVKISVNAEOVRTELEPPDAKPLGVQQRQKQMOOKSODLESVOEVG 1200  
DB 1141 SHQAQVTSYFVKISVNAEOVRTELEPPDAKPLGVQQRQKQMOOKSODLESVOEVG 1200  
QY 1201 GSYWQVRYTLLELLOHKKKLSPOILVPTLNLSSRLBLEPQEOGMEYTKOLISCLL 1260  
DB 1201 GSYWQVRYTLLELLOHKKKLSPOILVPTLNLSSRLBLEPQEOGMEYTKOLISCLL 1260  
QY 1261 NIOQKLSPPDGCKIPKOILDEKFNVELIVQCIRLSEMPQHNHALLLGLVAGIFPPKVL 1320  
DB 1261 NIOQKLSPPDGCKIPKOILDEKFNVELIVQCIRLSEMPQHNHALLLGLVAGIFPPKVL 1320  
QY 1321 HNIMSIPTFMGAVNMRLDPTYSFQVINKTVKAVYIPALIOGSDGSEIEVSANVEIIVKII 1380  
DB 1321 HNIMSIPTFMGAVNMRLDPTYSFQVINKTVKAVYIPALIOGSDGSEIEVSANVEIIVKII 1380  
QY 1381 SVFVDALPHVDEHRRRLPIVLQVDTLGAEXFLMILLLEFQYVTKTVLAAAYGEKDAIL 1440  
DB 1381 SVFVDALPHVDEHRRRLPIVLQVDTLGAEXFLMILLLEFQYVTKTVLAAAYGEKDAIL 1440

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:46:09 ; Search time 37.5044 Seconds  
(without alignments)  
5003.834 Million cell updates/sec

Title: US-09-603-665-5  
Perfect score: 10803  
Sequence: 1 MTSLAQQLQRLALPQSDASL.....CCKTIQQLQETVLGEPLQSYF 2144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /EMC\_Celerra\_SIDS3/prodata/2/1aa/5-COMB.pep.\*  
2: /EMC\_Celerra\_SIDS3/prodata/2/1aa/6-COMB.pep.\*  
3: /EMC\_Celerra\_SIDS3/prodata/2/1aa/7-COMB.pep.\*  
4: /EMC\_Celerra\_SIDS3/prodata/2/1aa/H-COMB.pep.\*  
5: /EMC\_Celerra\_SIDS3/prodata/2/1aa/PTCUTS-COMB.pep.\*  
6: /EMC\_Celerra\_SIDS3/prodata/2/1aa/RE-COMB.pep.\*  
7: /EMC\_Celerra\_SIDS3/prodata/2/1aa/backfilltest1.pep.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	427.5	4.0	721	2 US-09-248-796A-15107	Sequence 15107, A
2	310	2.9	200	2 US-09-248-796A-15108	Sequence 15108, A
3	272.5	2.5	2954	2 US-09-150-867-1	Sequence 1, Appli
4	272.5	2.5	2954	3 US-09-724-584-1	Sequence 1, Appli
5	268.5	2.5	2663	2 US-09-538-092-1352	Sequence 1252, Ap
6	266	2.5	2662	2 US-09-595-684B-311	Sequence 31, Appli
7	262	2.4	2516	2 US-09-949-016-10280	Sequence 10280, A
8	262	2.4	3056	1 US-08-508-836A-8	Sequence 8, Appli
9	262	2.4	3056	1 US-08-629-001A-3	Sequence 3, Appli
10	262	2.4	3056	1 US-08-874-266-2	Sequence 2, Appli
11	262	2.4	3056	2 US-08-642-274D-3	Sequence 3, Appli
12	262	2.4	3056	2 US-08-952-127-3	Sequence 3, Appli
13	262	2.4	3056	2 US-08-952-014C-3	Sequence 2, Appli
14	262	2.4	3056	2 US-08-984-090-2	Sequence 1, Appli
15	262	2.4	3056	2 US-10-351-733-1	Sequence 3, Appli
16	260.5	2.4	3057	2 US-09-360-416-3	Sequence 3, Appli
17	260.5	2.4	3057	2 US-10-175-225-3	Sequence 12, Appli
18	259	2.4	3066	2 US-08-952-127-12	Sequence 10, Appli
19	255.5	2.4	3433	2 US-09-091-501B-10	Sequence 1136, Ap
20	255.5	2.4	3433	2 US-09-538-092-1136	Sequence 2, Appli
21	255	2.4	3056	2 US-09-360-416-2	Sequence 2, Appli
22	255	2.4	3056	2 US-10-175-225-2	Sequence 2, Appli
23	248	2.3	1979	2 US-09-949-016-6468	Sequence 6468, Ap
24	248	2.3	2047	2 US-09-949-016-7404	Sequence 7404, Ap
25	247.5	2.3	10182	2 US-09-134-001C-3159	Sequence 3159, Ap
26	236.5	2.2	2470	2 US-08-265-967C-2	Sequence 2, Appli

27	236.5	2.2	2470	2 US-08-305-790B-3	Sequence 3, Appli
28	235	2.2	2482	1 US-08-328-254-6	Sequence 6, Appli
29	225.5	2.1	3259	2 US-09-949-016-6507	Sequence 6507, Ap
30	224.5	2.1	3210	2 US-09-538-092-1154	Sequence 1154, Ap
31	223	2.1	3830	2 US-09-693-205A-4	Sequence 4, Appli
32	222.5	2.1	3248	1 US-08-353-700-1	Sequence 1, Appli
33	222.5	2.1	3248	5 PCT-US95-16216-1	Sequence 16, Appli
34	220	2.0	3829	2 US-09-693-205A-16	Sequence 2, Appli
35	218.5	2.0	1972	2 US-08-875-435B-3	Sequence 3, Appli
36	218	2.0	3829	2 US-09-693-205A-2	Sequence 11433, A
37	216.5	2.0	2733	2 US-09-949-016-11433	Sequence 11, Appli
38	216	2.0	3878	2 US-09-914-258-11	Sequence 87, Appli
39	214.5	2.0	2108	2 US-09-538-092-87	Sequence 1077, Ap
40	214	2.0	1960	2 US-09-538-092-1077	Sequence 10872, A
41	213	2.0	1960	2 US-09-949-016-10872	Sequence 34, Appli
42	209	1.9	4096	2 US-09-296-662-34	Sequence 1903, Ap
43	207.5	1.9	2209	2 US-10-017-754-1903	Sequence 8, Appli
44	207	1.9	1786	2 US-08-973-462-8	Sequence 4, Appli
45	206.5	1.9	1972	2 US-08-875-435B-4	

ALIGNMENTS

RESULT 1  
US-09-248-796A-15107  
Sequence 15107, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Kelch weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/1248, 796A  
PRIOR FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074, 725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096, 409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ. ID NOS: 28208  
SEQ ID NO 15107  
LENGTH: 721  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15107  
Query Match 4.0%; Score 427.5; DB 2; Length 721;  
Best Local Similarity 22.3%; Pred. No. 3.7e-26;  
Matches 175; Conservative 152; Mismatches 311; Indels 145; Gaps 28;  
QY 1436 KDALEADTEFWFVSCFVSQVQHIOQLMNILOYILKLPKKEETIPKAVSFNKSESOE 1495  
DB 7 KNQDSQVAFVFGVYGMKSFSAEBQLAGIVAFTKLMNDIP-----LNQLEPGE 54  
QY 1496 MLQFN-----VEHTSKQLRHFKFLSVSPQSLSNN-----FLKK-----VYES 1537  
DB 55 EFVELNRRPVGCTTATLGGSGGLAVLRNDLQFLDETLSSBNKHELSLTKRMALVLIDD 114  
QY 1538 GGPETLKGLEERLLETLYGVISAVASQSWERNADKLT-----VKFWRALSLKAYDLIDKVN 1592  
DB 115 EDSEVSK-----KELVLDKRRATITFPALSLDFTFNSHAIK-----LCSTLYSLGNML 164  
QY 1593 ALPTEFPI-PVINGLVGNPLP-----SVRRKALDLNNKLAQ--NISKWTIVTRFKL 1644  
DB 165 DLPPLNYTIDISVSLDVTLSDSIKVARNVAILASRKEETELNVACHQVVE--SV 222  
QY 1645 VPDLLAVQRKKKGEEEQALNRRQALTYLTKLCNFGAENDP-----PFVPLXTAVKLI 1700  
DB 223 INNLPLIKIGIKKKNVDVL--QQAVLDFTSTIVNKKFASGTFPLSDVSKLISLGIV 280  
QY 1701 APER---KEKNVIGSALLCIAEVTSTLEALAIPLPSLMBSLTTMKNTSELVSSEVYL 1757  
DB 281 TTDRGLNLEQPEVITIASINAIITSIVNIGVKTLGLFPKVPFPAKIMWESTNLSGDKESAK 340

QY 1756 L---SALAAQGVVELLPFISPYLEGILSQVYIHLEKXTSENGSA----- 1799  
Db 341 LAGGSYLVLLSCYIKKIPAFMSTTLEAVLLTTLSSDLDNHRRSSVLDLVDHMDLAQVL 400  
QY 1800 -SQANIRLTSLKKTUATTAPRVLLPAIKKTYKOIEKMKKNHMGPFMSILOEHIGMKKE 1856  
Db 401 KSLCNVWL-----KKFV---TNDNSGNIIGFLTKTQATINMKKK 438  
QY 1859 ELTSHOSQITAEFLALDIPRAQHSN-----DLEEVGKTENCIDICLVAMVYKSEVTR 1913  
Db 439 QATTQATLEMRWLISAFEEF-QYSEEDNKKFPDNNTIHRLSESFHCALAFWKLKDKSFR 497  
QY 1914 PLFFFLFDMA-----KTEDAPKDRLLTFENLADCIAEKKGFTLEAGHLYVFPADT 1965  
Db 498 PLFANLVRAVDGEGATLKTNEV--SRLLAFRRFRKQDELKSLTTSYFSLDLPJAL 555  
QY 1966 LKQVNIKSTDEAFPSSENDPEKCCLLQFLNLCKYKFLFTDHFISKE-BAKALIMBLV 2024  
Db 556 L-----KPFSESLVATN-----LRILLGLTSSFFYDDDDVWSQGGFSDJCSPLL 603  
QY 2025 DQENRLOGEEKFQRRVTKHLIPCIAGPSVAMADSLMKPLNYQI--LKTROSSPKVR 2081  
Db 604 SOLSN-----IEDSIGKYLKVSSTFVTQVDSDEYMETVLVHELKYSINANENSATK 656  
QY 2082 FALITVTLALAEKLENYIVLLPESIPFLAIMEDECEVEHQQCK-TIQOLETVLGEPL 2140  
Db 657 IWSITLTKTIFQKMGQWLSYLPYIAELLEDDEBEVEMVARGLVATENVLGEPL 716  
QY 2141 QSY 2143  
Db 717 DRY 719  
RESULT 2  
US-09-248-796A-15108  
; Sequence 15108, Application US/09248796A  
; Patent No. 6747137  
GENERAL INFORMATION :  
APPLICANT: Keith Weinstein et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 15108  
LENGTH: 200  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15108  
Query Match 2.9%; Score 310; DB 2; Length 200;  
Best Local Similarity 38.0%; Pred. No. 2.4e-17;  
Matches 76; Conservative 36; Mismatches 70; Indels 18; Gaps 5;  
QY 1 MTSLAQQLQRL-----ALPQSDASLL-SRDEVASLLFDPKEAATIDRDTAFAGTGL 52  
Db 4 MSSLSAQQSQISINEKTASVALDRKQSKLHSR---SLTFDPKQATQCYEYIEIATBGL 59  
QY 53 EELLGIDPSFEQFAPLPSQAKTIERSVQTAANKQDENISLPLHLSYFLLKPAQK 112  
Db 60 EDLCGLDRFNFKFLTLFSETSVNLDRNLQTDVVISQDKNDAFLTLVGGYGLTSSLK 119  
QY 113 CLEWLIHFPHILYKQDSLIACVLRYNHETRIFFRVYIOLKINNSGHRFWMLLPVK-QSGV 171  
Db 120 AVENLVRRFPHANIHAEMLITLALPYQGRVVEVAVLVNIKRONLPQIEWLVGKQDLKT 179  
QY 172 PLAKGTLLTHCYKDLGFMDF 191

```

Db      180 PPA$IL-----KGM$AMNF 194

RESULT 3
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowitz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-directed Microtubule Motor Required for
; FILE REFERENCE: 1957C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2954
; TYPE: PRT
; ORGANISM: Xenopus sp.
; FEATURE:
; OTHER INFORMATION: Xenopus centromere-associated protein-E (XCENP-E)
; OTHER INFORMATION: member of the kinesin superfamily of microtubule
; OTHER INFORMATION: motor proteins
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1)..(472)
; OTHER INFORMATION: kinesin like motor domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (473)..(2752)
; OTHER INFORMATION: rod domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (2753)..(2954)
; OTHER INFORMATION: tail domain
US-09-150-867-1

Query Match      2.5% Score 272.5; DB 2; Length 2954;
Best Local Similarity 18.7%; Pred. No. 3,1e-12;
Matches 415; Conservative 365; Mismatches 809; Indels 629; Gaps 96.

1 MTSLAQQLQRLALPQSDASLISRDENVASLLFDPKKAATIDRDYFAIGCTGLEILLGDP 60
Db      702 MAEKANLLELA-----LMRDNPNIIL--ENETLKREILA-----DLERSLKENQ 744
QY      61 SFDEFEAPLPSQLAKTIERSVQTKAVN-----KQIDENISLFLHLSPYF-----LLKPA 110
Db      745 ETNFEET---LEKETQKEHQAQILHEIGSLKGLVNAEMYNQNLBEDLETQTKLKEQ 799
QY      111 QKCLEWL-----IHREHILYNQDSLIIACVLPYHETRIFFRVYIQLIKINNSKRW 160
Db      800 EIQLAELRKRADNLQKKVRNPDLSVSGDSKLC-----EEIFQL-----839
QY      161 FWLLPVGQCVPLAKGLLTH-CYKDLGFM-----DFICSLVTK 198
Db      840 -----KQS---LSDAEAVTRDAQKESFLSENLLEKEMEDTSMNYNQEKASLFEK 890
QY      199 SVKV-PAEYPPSSAQLRVLL-AFYASTIVSALVAEDVSQNI---IAKLPYIQGLK 251
Db      891 QLETERKSNYKMGKADLQKELOSAFENEINYINGLAGKVPRLDLRVELEKQVSEFSKQLE 950
QY      252 SSLPDYAAATYMTICQISVAYTMENTVNSLASQIITLTFTPLSIDGLSGCLVLLQRO 311
Db      951 KALEEKALENEVCLESEYKFLPNE--VECLKNDISKASEIIMLLKQEGHSASII---S 1005

```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:52:34 ; Search time 134.107 Seconds  
(without alignments)  
7405.549 Million cell updates/sec

Title: US-09-603-665-5  
Perfect score: 10803  
Sequence: 1 MTSLAQOLRLALPQSDASL.....CCKTIQQLTVLGEPLQSYF 2144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /EMC\_Celera\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celera\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celera\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celera\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 5: /EMC\_Celera\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celera\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10799	99.9	2144	5	US-10-745-237-230 Sequence 230, App
2	10782	99.8	2144	6	US-11-124-368A-277 Sequence 277, App
3	10722	94.6	2044	6	US-11-124-368A-278 Sequence 278, App
4	9853	91.2	2036	6	US-11-124-368A-276 Sequence 276, App
5	9853	91.2	2036	6	US-11-124-368A-280 Sequence 280, App
6	9853	91.2	2036	6	US-11-124-368A-281 Sequence 281, App
7	7756.5	71.8	1569	5	US-10-450-763-45591 Sequence 45591, A
8	5026.5	46.5	1149	5	US-10-128-558-167 Sequence 167, App
9	4494	41.6	897	4	US-10-764-425-177 Sequence 177, App
10	3694	34.2	734	4	US-10-108-260A-3300 Sequence 3300, App
11	2061	19.1	408	6	US-11-124-368A-279 Sequence 279, App
12	2051.5	19.0	2096	5	US-11-124-368A-277 Sequence 277, App
13	2051.5	19.0	2096	5	US-11-097-143-22518 Sequence 36, Appl
14	1603	14.8	325	5	US-10-128-558-358 Sequence 358, App
15	1169	10.8	2122	4	US-10-437-963-189782 Sequence 189782, A
16	1086	10.1	1798	4	US-10-128-714-8388 Sequence 8388, App
17	1085	10.0	1814	4	US-10-128-714-8388 Sequence 8388, App
18	1056.5	9.8	1818	4	US-10-032-585-7616 Sequence 7616, App
19	673	6.2	952	4	US-10-424-599-24811 Sequence 24811, A
20	663	6.1	135	4	US-10-264-049-2412 Sequence 2412, App
21	512.5	4.7	463	4	US-10-425-115-295784 Sequence 295784, A
22	434	4.0	437	4	US-10-425-114-44447 Sequence 44447, A
23	379	3.5	77	3	US-09-864-761-37036 Sequence 37036, A
24	309.5	2.9	2228	5	US-10-511-096-2 Sequence 2, Appl
25	309.5	2.9	2230	5	US-10-511-096-4 Sequence 4, Appl
26	308	2.9	2250	5	US-10-511-096-6 Sequence 6, Appl
27	308	2.9	2252	5	US-10-511-096-8 Sequence 8, Appl

28	277	2.6	2668	6	US-11-124-368A-215 Sequence 215, App
29	275.5	2.6	2633	5	US-10-450-763-36864 Sequence 36864, A
30	274	2.5	2834	4	US-10-424-599-256711 Sequence 256711, A
31	273.5	2.5	2665	6	US-11-124-368A-214 Sequence 214, App
32	272.5	2.5	2954	5	US-10-650-280-1 Sequence 1, Appl
33	268.5	2.5	2663	5	US-10-723-860-749 Sequence 749, App
34	268	2.5	2543	5	US-10-828-985A-9 Sequence 9, Appl
35	267.5	2.5	2733	6	US-11-097-143-9003 Sequence 9003, App
36	265.5	2.5	2405	4	US-10-437-963-106815 Sequence 106815, App
37	265	2.5	2503	5	US-10-828-985A-11 Sequence 11, Appl
38	264.5	2.4	2568	5	US-10-828-985A-7 Sequence 7, Appl
39	262	2.4	3056	2	US-08-984-090-2 Sequence 2, Appl
40	262	2.4	3056	3	US-09-825-476-3 Sequence 3, Appl
41	262	2.4	3056	4	US-10-060-065-20 Sequence 20, Appl
42	262	2.4	3056	4	US-10-307-077-1 Sequence 1, Appl
43	262	2.4	3056	4	US-10-059-585-41 Sequence 41, Appl
44	262	2.4	3056	4	US-10-361-562-2 Sequence 2, Appl
45	262	2.4	3056	4	US-10-116-275-338 Sequence 338, App

## ALIGNMENTS

RESULT 1  
US-10-745-237-230  
Sequence 230, Application US/10745237  
Publication No. US20050227301A1  
GENERAL INFORMATION:  
APPLICANT: Cyclacel Limited  
APPLICANT: Glover, David  
APPLICANT: Bell, Graham  
APPLICANT: Frenz, Lisa  
APPLICANT: Midgley, Carol  
TITLE OF INVENTION: Cell Cycle Progression Proteins  
FILE REFERENCE: P015693MO CYK  
CURRENT APPLICATION NUMBER: US/10/745,237  
CURRENT FILING DATE: 2003-12-23  
PRIOR APPLICATION NUMBER: US 60/439,123  
PRIOR FILING DATE: 2003-01-10  
PRIOR APPLICATION NUMBER: US 60/468,402  
PRIOR FILING DATE: 2003-05-06  
NUMBER OF SEQ ID NOS: 600  
SOFTWARE: Patent version 3.1  
SEQ ID NO 230  
LENGTH: 2144  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Q9H563  
US-10-745-237-230

Query Match 99.9%; Score 10799; DB 5; Length 2144;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	MTSLAQOLRLALPQSDASLISRDVALLFDPRKAAATIDRTAFAGCTGLEBLLGDP	60
DB	1	MTSLAQOLRLALPQSDASLISRDVALLFDPRKAAATIDRTAFAGCTGLEBLLGDP	60
QY	61	SFEQFEPPLSQLAKTERSVQTKAVNKQDENISLPLHLSPYFLPKPAQKCLEMLIHR	120
DB	61	SFEQFEPPLSQLAKTERSVQTKAVNKQDENISLPLHLSPYFLPKPAQKCLEMLIHR	120
QY	121	FHILVNQDSLILACVLPYHETRIFFVYIQLKINNSGHRFWMLLPVQSGVPLAKGLTIT	180
DB	121	FHILVNQDSLILACVLPYHETRIFFVYIQLKINNSGHRFWMLLPVQSGVPLAKGLTIT	180
QY	181	HCYKXGFMDFISLVTKSVKFAEYIPGSSAQRLVLLAFYASTIVSALVAEDVSDNIIA	240
DB	181	HCYKXGFMDFISLVTKSVKFAEYIPGSSAQRLVLLAFYASTIVSALVAEDVSDNIIA	240
QY	241	KLEPYIOXGKSLSPYRATVMIICQISVKYKMENTFNUSLNSQITKTKPSLTKDG	300
DB	241	KLEPYIOXGKSLSPYRATVMIICQISVKYKMENTFNUSLNSQITKTKPSLTKDG	300

Db	241	KLFPYIQKGLKSSLPYRAATWMIICQISVKTMENTFVNSLASQITIKTLKIBELMDG	300
Qy	301	LSCLIVALLQCKEBSLQKPPFPLCNVDPDLITLHGISETYDVSPLLRMYLPHLVSIH	360
Db	301	LSCLIVALLQCKEBSLQKPPFPLCNVDPDLITLHGISETYDVSPLLRMYLPHLVSIH	360
Qy	361	HVTGEETGMDGOQYKRLHFAILTKISLKNLDHLLASLLFEEYISYSYSGEEMDNKXSL	420
Db	361	HVTGEETGMDGOQYKRLHFAILTKISLKNLDHLLASLLFEEYISYSYSGEEMDNKXSL	420
Qy	421	LNBOFLPLILBESKYPRTLDVLYEHLXELADLKXOEFLPHQVSLSTSGKXQFLAOSD	480
Db	421	LNBOFLPLILBESKYPRTLDVLYEHLXELADLKXOEFLPHQVSLSTSGKXQFLAOSD	480
Qy	481	TSMLSLNHP LAPRIILAMNHLKKIMTSGEGDESFKEAVLARIGDNDIVLSAIS	540
Db	481	TSMLSLNHP LAPRIILAMNHLKKIMTSGEGDESFKEAVLARIGDNDIVLSAIS	540
Qy	541	FEIFKEHFSSEVTISNLTNLFORALSKNGEYEVLYKIAADILKEIILSENDOLSNQV	600
Db	541	FEIFKEHFSSEVTISNLTNLFORALSKNGEYEVLYKIAADILKEIILSENDOLSNQV	600
Qy	601	VCLLPFVYVINDDTESAEKAIATLSGSGCSLPHLRGWEALENVISTRGKLGIVA	660
Db	601	VCLLPFVYVINDDTESAEKAIATLSGSGCSLPHLRGWEALENVISTRGKLGIVA	660
Qy	661	NOKMIELLADININGDPSSMLKXWEDLISYGEESFNLKOKVFNHILSVYSCSSLKE	720
Db	661	NOKMIELLADININGDPSSMLKXWEDLISYGEESFNLKOKVFNHILSVYSCSSLKE	720
Qy	721	THFPFAIRVFSLLQKKIKLESVITTAVEIPSEWHIELMDRGIPVELMAYVEELNSTOR	780
Db	721	THFPFAIRVFSLLQKKIKLESVITTAVEIPSEWHIELMDRGIPVELMAYVEELNSTOR	780
Qy	781	VAVEDSVLVYLSIKKETIYALKAPSPFGDIWMNPEQKEDSRDYHLILGLPEMMINGA	840
Db	781	VAVEDSVLVYLSIKKETIYALKAPSPFGDIWMNPEQKEDSRDYHLILGLPEMMINGA	840
Qy	841	DAVHFRVLMKLFIKYHLEDVQFLFKFCVLTWTGSSLSNPLNCSVKTVLOTQALYGCAM	900
Db	841	DAVHFRVLMKLFIKYHLEDVQFLFKFCVLTWTGSSLSNPLNCSVKTVLOTQALYGCAM	900
Qy	901	LSQKTOCKHQLASISSPVYVTSLLINLCSPVKEVRRAAIOCLQALSGVAPFYILIDHLI	960
Db	901	LSQKTOCKHQLASISSPVYVTSLLINLCSPVKEVRRAAIOCLQALSGVAPFYILIDHLI	960
Qy	961	SKAEETSPAAYVIODLATLFEELOREKKLSHOKLSTLKNLSCVYSCSYIAXDKLMK	1020
Db	961	SKAEETSPAAYVIODLATLFEELOREKKLSHOKLSTLKNLSCVYSCSYIAXDKLMK	1020
Qy	1021	VLOGNVGENVLSQLPMAEQULEKIQKEPTAVLKDENVLHLTLGKYNEFSVSLNEDPK	1080
Db	1021	VLOGNVGENVLSQLPMAEQULEKIQKEPTAVLKDENVLHLTLGKYNEFSVSLNEDPK	1080
Qy	1081	SLDIFIKAVHTTKELYAGMPTIOITALEKITRPFPAAISDEKVQKILRMFLDLVANCN	1140
Db	1081	SLDIFIKAVHTTKELYAGMPTIOITALEKITRPFPAAISDEKVQKILRMFLDLVANCN	1140
Qy	1141	SHCAQVYVSSFKGISYMAQVRIELEBPPKAPPLGIVQOQKROKQOQKKSODLSVEVGG	1200
Db	1141	SHCAQVYVSSFKGISYMAQVRIELEBPPKAPPLGIVQOQKROKQOQKKSODLSVEVGG	1200
Qy	1201	GSYVORVTLTLELLOHKKKLSRPOILVPTLFNLISHCLEPLREBOGQNNETKOLILSCL	1260
Db	1201	GSYVORVTLTLELLOHKKKLSRPOILVPTLFNLISHCLEPLREBOGQNNETKOLILSCL	1260
Qy	1261	NIQOKLSPDGSKI PKOILDEEKENVELIYOCIRLSMPOTHHNALLILGTAVGIFPDKVL	1320
Db	1261	NIQOKLSPDGSKI PKOILDEEKENVELIYOCIRLSMPOTHHNALLILGTAVGIFPDKVL	1320
Qy	1321	HNIMSIITPFGANVNRDLDTYOSOVINKYVAVIPALIOSDSDGSI EYSRNVEEIVXKI	1380
Db	1321	HNIMSIITPFGANVNRDLDTYOSOVINKYVAVIPALIOSDSDGSI EYSRNVEEIVXKI	1380

QY	1381	SVFVDALPHVEHRRRLPIVLQVDTLGAENELMILLILFEQVYTKVLAAYGEXKAIL	1440
Db	1381	SVFVDALPHVEHRRRLPIVLQVDTLGAENELMILLILFEQVYTKVLAAYGEXKAIL	1440
QY	1441	EADTEPFMSVCCFESVOHQIOSLMNIILOYLLKLPEBEKETPKAVSEFKSSQOEMLQVF	1500
Db	1441	EADTEPFMSVCCFESVOHQIOSLMNIILOYLLKLPEBEKETPKAVSEFKSSQOEMLQVF	1500
QY	1501	NVEHTHSQQLRHPFPLASVFMSSOLLSSNNPKKVVESGPEILKGLERRLLETVLGYSIA	1560
Db	1501	NVEHTHSQQLRHPFPLASVFMSSOLLSSNNPKKVVESGPEILKGLERRLLETVLGYSIA	1560
QY	1561	VAOSMERADKLTVKFWRALSKAYDLLDKYNALLPTEFIPVIRGLVGNPLPSVRKAL	1620
Db	1561	VAOSMERADKLTVKFWRALSKAYDLLDKYNALLPTEFIPVIRGLVGNPLPSVRKAL	1620
QY	1621	DLLNNKLOONISWKKTIYTRFLKLVDPDLAIYQRKKXGEEBOALNROFALVTLLKLCGN	1680
Db	1621	DLLNNKLOONISWKKTIYTRFLKLVDPDLAIYQRKKXGEEBOALNROFALVTLLKLCGN	1680
QY	1681	FAENRDPFVPLTXAVKLIAPERKEEKVLSALICIAEVSTLEALAIPLPLPSLMPSL	1740
Db	1681	FAENRDPFVPLTXAVKLIAPERKEEKVLSALICIAEVSTLEALAIPLPLPSLMPSL	1740
QY	1741	LTTMNTSELVSESVYLLSALAALQKVETLPHFISPYLEGILSOVILHEKITSEMGAS	1800
Db	1741	LTTMNTSELVSESVYLLSALAALQKVETLPHFISPYLEGILSOVILHEKITSEMGAS	1800
QY	1801	QANIRLTSIKTLATTLAPRVLLPAIKTYKQIEKRWKXHMGPMSILDEHIGXMKKEEL	1860
Db	1801	QANIRLTSIKTLATTLAPRVLLPAIKTYKQIEKRWKXHMGPMSILDEHIGXMKKEEL	1860
QY	1861	TSHQSLTFPLFALDPFAQHSSENDLEEVGKTENCIIICLVAMVVKLSEVTFRPLPFKLF	1920
Db	1861	TSHQSLTFPLFALDPFAQHSSENDLEEVGKTENCIIICLVAMVVKLSEVTFRPLPFKLF	1920
QY	1921	DWAKTEDAPKDRLLTFYNLADICIAEKLKGLFTLFAHLVYKFPADTLXOVNISKTDEAFD	1980
Db	1921	DWAKTEDAPKDRLLTFYNLADICIAEKLKGLFTLFAHLVYKFPADTLXOVNISKTDEAFD	1980
QY	1981	SENDPEKCCLLLOFLINCLYKIFLPPTQHFISKERAXALMMPVDOLENLGGEEKFOER	2040
Db	1981	SENDPEKCCLLLOFLINCLYKIFLPPTQHFISKERAXALMMPVDOLENLGGEEKFOER	2040
QY	2041	VTHGLPLCIAQFSVAMADSLMKPLNVOILLKTRDSSPKVRFAALLTVLALAEKUKENYI	2100
Db	2041	VTHGLPLCIAQFSVAMADSLMKPLNVOILLKTRDSSPKVRFAALLTVLALAEKUKENYI	2100
QY	2101	VLPESIPFLAELMEDECEVEHQCCQTTIOLETVLGEPIQSYF	2144
Db	2101	VLPESIPFLAELMEDECEVEHQCCQTTIOLETVLGEPIQSYF	2144
RESULT 2			
US-11-124-368A-277			
Sequence 277, Application US/11124368A			
Publication No. US2005028759A1			
GENERAL INFORMATION:			
APPLICANT: Michele Cargill			
APPLICANT: James J. Devlin			
APPLICANT: May Luke			
TITLE OF INVENTION: Genetic Polymorphisms Associated with			
TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof			
FILE REFERENCE: C1001524			
CURRENT APPLICATION NUMBER: US/11/124,368A			
CURRENT FILING DATE: 2005-05-09			
PRIOR APPLICATION NUMBER: US 60/568,845			
PRIOR FILING DATE: 2004-05-07			
PRIOR APPLICATION NUMBER: US 60/625,936			
PRIOR FILING DATE: 2004-11-09			
NUMBER OF SEQ ID NOS: 21112			
SOFTWARE: fastseq for windows Version 4.0			



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2006, 00:53:49 ; Search time 11.9332 Seconds

(without alignments)  
4243.670 Million cell updates/sec

Title: US-09-603-665-5

Perfect score: 10803

Sequence: 1 MTSLAQOLQRLALPQSDASL.....CCKTIOQLTVGEPLQSYF 2144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 103426 segs, 23619683 residues

Total number of hits satisfying chosen parameters: 103426

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3694	34.2	734	7	US-11-293-697-3300 Sequence 3300, App
2	235	2.2	3113	6	US-10-505-928-325 Sequence 325, App
3	207.5	1.9	2209	7	US-11-301-554-1903 Sequence 1903, App
4	205	1.9	93	6	US-10-449-902-34905 Sequence 34905, A
5	205	1.9	93	6	US-10-449-902-35765 Sequence 35765, A
6	188.5	1.7	1328	6	US-10-504-973-332 Sequence 32, App1
7	184.5	1.7	9635	6	US-10-441-571A-4496 Sequence 4496, App
8	179	1.7	1206	6	US-10-449-902-43305 Sequence 43305, A
9	175	1.6	1205	7	US-11-293-697-2367 Sequence 2367, App
10	175	1.6	1842	6	US-10-511-937-2929 Sequence 2929, App
11	173.5	1.6	1176	6	US-10-449-902-46322 Sequence 46322, A
12	171	1.6	1325	6	US-10-449-902-51794 Sequence 51794, A
13	166	1.5	1346	6	US-10-449-902-55021 Sequence 55021, A
14	163.5	1.5	2871	6	US-10-505-928-100 Sequence 100, App
15	161	1.5	1218	6	US-10-449-902-47146 Sequence 47146, A
16	159.5	1.5	2671	6	US-10-505-928-784 Sequence 784, App
17	158.5	1.5	1116	6	US-10-449-902-47106 Sequence 47106, A
18	157.5	1.5	980	7	US-11-289-102-296 Sequence 296, App
19	155	1.4	764	6	US-10-449-902-45961 Sequence 45961, A
20	152.5	1.4	1230	7	US-11-289-102-378 Sequence 378, App
21	152	1.4	748	6	US-10-449-902-53845 Sequence 53845, A
22	152	1.4	2364	7	US-11-289-102-242 Sequence 242, App
23	151.5	1.4	1368	7	US-11-283-329-166 Sequence 166, App
24	150.5	1.4	1120	6	US-10-505-928-213 Sequence 213, App
25	150	1.4	1050	6	US-10-471-571A-4624 Sequence 4624, App

26	149	1.4	1443	6	US-10-486-020-14 Sequence 14, App1
27	148.5	1.4	868	6	US-10-953-349-674 Sequence 674, App
28	148.5	1.4	870	6	US-10-953-349-673 Sequence 673, App
29	148.5	1.4	1217	6	US-10-471-571A-4942 Sequence 4942, App
30	147.5	1.4	868	6	US-10-449-902-43523 Sequence 43523, A
31	146.5	1.4	784	6	US-10-953-349-675 Sequence 675, App
32	146.5	1.4	1146	6	US-10-471-571A-4478 Sequence 4478, App
33	145.5	1.3	1056	6	US-10-505-928-225 Sequence 225, App
34	144.5	1.3	1114	6	US-10-449-902-52920 Sequence 52920, A
35	144.5	1.3	1315	6	US-10-527-411-141 Sequence 141, App
36	143	1.3	775	7	US-11-293-697-2844 Sequence 2844, App
37	142.5	1.3	1114	6	US-10-449-902-46281 Sequence 46281, A
38	142.5	1.3	1663	6	US-10-505-928-227 Sequence 227, App
39	141.5	1.3	1056	6	US-10-449-902-45516 Sequence 45516, A
40	141	1.3	1575	6	US-10-505-928-257 Sequence 257, App
41	140.5	1.3	742	6	US-10-953-349-11696 Sequence 11696, A
42	139.5	1.3	1174	6	US-10-449-902-49996 Sequence 49996, A
43	139.5	1.3	1346	6	US-10-449-902-55428 Sequence 55428, A
44	139	1.3	990	7	US-11-293-697-3896 Sequence 3896, App
45	135.5	1.3	838	6	US-10-449-902-45252 Sequence 45252, A

ALIGNMENTS

RESULT 1  
US-11-293-697-3300 Application US/11293697  
; Sequence 3300, App1  
; Publication No. US200601053176A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length CDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3300  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-293-697-3300

Query Match 34.2% ; Score 3694 ; DB 7 ; Length 734 ;  
Best Local Similarity 100.0% ; Pred. No. 7.5e-206 ;  
Matches 734 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;

QY	1	MTSLAQOLQRLALPQSDASLSRDEVASILPDKKAAITDRDTAFAIGTGHEELLGIDP	60
DB	1	MTSLAQOLQRLALPQSDASLSRDEVASILPDKKAAITDRDTAFAIGTGHEELLGIDP	60
QY	61	SPEQEAFLPQSALTLERSVOTKAVNKQDENISLPLIHSPLYLKPQCKLEMLIHR	120
DB	61	SPEQEAFLPQSALTLERSVOTKAVNKQDENISLPLIHSPLYLKPQCKLEMLIHR	120
QY	121	FHILYNDSLIACVLPYHETRIFFRVITQLKINNSKIRFWMLPVKOSGVPLAGTLIT	180
DB	121	FHILYNDSLIACVLPYHETRIFFRVITQLKINNSKIRFWMLPVKOSGVPLAGTLIT	180
QY	181	HCKYKLGFMDFCSIWTSKVVAFAEYPSASQRLVLAIFYASTISALVAADVDNIIA	240
DB	181	HCKYKLGFMDFCSIWTSKVVAFAEYPSASQRLVLAIFYASTISALVAADVDNIIA	240
QY	241	KLFPIYQKGLKSSLPDYRAATYMIICQISVKYTMENFVNSLASQIIKTLTKIPSLIKDG	300
DB	241	KLFPIYQKGLKSSLPDYRAATYMIICQISVKYTMENFVNSLASQIIKTLTKIPSLIKDG	300
QY	301	LSCLIVLLQKQKPSLGGKPPHLCNVVDLITLIGISFTYVSLPLRLMPLHVLVSIH	360
DB	301	LSCLIVLLQKQKPSLGGKPPHLCNVVDLITLIGISFTYVSLPLRLMPLHVLVSIH	360



QY 361 HTGTEEGMDGOIYKRNHLEAILTKISLKNNDHLLASLLFEEXISYSQGEENDSNKVSLL 420  
 DB 361 HTGTEEGMDGOIYKRNHLEAILTKISLKNNDHLLASLLFEEXISYSQGEENDSNKVSLL 420  
 QY 421 LNFQPLPLRLLESKYRPTLDVLEELKEIADLKKQELFHOFVSLSTSGGKYOPLADSD 480  
 DB 421 LNFQPLPLRLLESKYRPTLDVLEELKEIADLKKQELFHOFVSLSTSGGKYOPLADSD 480  
 QY 481 TSLMLSLNHLPLAVRIILANNHLLKIMKTSKEGVDESFIKEAVLARIJDDNIDVLSAISA 540  
 DB 481 TSLMLSLNHLPLAVRIILANNHLLKIMKTSKEGVDESFIKEAVLARIJDDNIDVLSAISA 540  
 QY 541 FEIFKEHFSSEVTISNLLNFORAEISKNGBYEVLKIAADILIKEIILENDQLSNOVV 600  
 DB 541 FEIFKEHFSSEVTISNLLNFORAEISKNGBYEVLKIAADILIKEIILENDQLSNOVV 600  
 QY 601 VCLLPVAVINDDTESAEKIAIYLSKSGICSLHPLRGWEALENVIKSTKPGKLIIVA 660  
 DB 601 VCLLPVAVINDDTESAEKIAIYLSKSGICSLHPLRGWEALENVIKSTKPGKLIIVA 660  
 QY 661 NQKMIELADNINLGPSSMLKMWEDLISVGEESFNLKOKVTFFHVLISVAVSCSSLKE 720  
 DB 661 NQKMIELADNINLGPSSMLKMWEDLISVGEESFNLKOKVTFFHVLISVAVSCSSLKE 720  
 QY 721 THFPFAIRVPSLLO 734  
 DB 721 THFPFAIRVPSLLO 734  
 RESULT 2  
 US-10-505-928-325  
 Sequence 325, Application US/10505928  
 Publication No. US20060088532A1  
 GENERAL INFORMATION:  
 APPLICANT: Ludwig Institute for Cancer Research et al.  
 TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
 FILE REFERENCE: 28967/39178  
 CURRENT APPLICATION NUMBER: US/10/505,928  
 PRIOR FILING DATE: 2004-08-27  
 PRIOR APPLICATION NUMBER: US 60/363,019  
 PRIOR FILING DATE: 2002-03-07  
 NUMBER OF SEQ ID NOS: 866  
 SOFTWARE: PatentIn 3.2  
 SEQ ID NO 325  
 LENGTH: 3113  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-505-928-325  
 Query Match 2.28; Score 235; DB 6; Length 3113;  
 Best Local Similarity 19.08; Pred. No: 6e-05;  
 Matches 427; Conservative 357; Mismatches 817; Indels 646; Gaps 101;  
 QY 182 CYKDLGMDPFCGLVTKSVKPAEYPCSSAQLRVILAFYASTIVSALVAEDSDNIIAK 241  
 DB 751 CYODL-----HAFY-----ESLRDL-----KSDASLVTHMEDHORSILA 785  
 QY 242 LPFYIYK-----GLKSLIPDYRAATYMIICQISVKTWENTFVNSLASQIIKTL 290  
 DB 786 ---FDQOPAHNSFANIGSGSWPSESE---CLLEADQSPKNS----- 824  
 QY 291 TKRPSLKLKSLCLVILQOKP---ESLGGKPPHLCNVPLDITLHIGISE----- 339  
 DB 825 ---AIIQNVDSLEFLESQOKONSQLOKO---CE---ELVOIKGEIEENIMKAEQMH 873  
 QY 340 -----TYVSPPLRYMLPHLVSIHHVGTERTEGMDGOIYKRNHLEAILTKISLKNL 392  
 DB 874 QSTVAETSGKISKLQEDTSAHQNV-VAETLSALENKKELQLNDKXETQAEIQELKKS 932  
 QY 393 DILLASLTFE-----EYISYSQGEENDS---NKVSL-----LNEQFLP 427  
 DB 933 NHHLEDSLKEIQLISETLSP-EKKEMSGIISLNKREIEELTQENGTLKEINASLNEKKN 991

QY 428 LIRLESKYRPTLDVLEELKEIADLKKQELFHOFVSLSTSGGKYOPLA-----DSD 480  
 DB 992 LIOKSES-FANYID-EREKSISELSDYOKEKTLILORCEEFGNAYEDLSQRYKAAOEKN 1049  
 QY 481 TSLMLSLNHLPLAVRIILANNHLLKIMKTSKEGVDESFIKE-----AVLARIJDDNIDVLS 535  
 DB 1050 SKJECCLN-----CTSLCENRKELEBOLKEAFKHOEFLTKLFAFERNNQML 1100  
 QY 536 SAISAFELFEHFSSEVTISNLLNFORAEISKNGBYEVLKIAADILIKEIILS--EN 592  
 DB 1101 E-----LEIYQALRSEBMTDN-----QNNKSEAG-----LKQETMLTKERO 1138  
 QY 593 DQLSNOVVCLLPVAVINDDTESAEKIAIYLSKSGICSLHPLRGWEALENVIKSTK 652  
 DB 1139 NMQKQEV-----NDLLQENQOLMKVKTGRCONLE-----SEPIRNSKERE 1181  
 QY 653 PGKLGAVANO-----KMIELADNINLGPSSMLKMWEDLISVGEESFNLKOKVTFFHVLISVAVSCSSLKE 686  
 DB 1182 SER-----NQCFKQMDLEVAEISLDSYNADVLEAMLRNKEIKLQESKEKECLOHE 1236  
 QY 687 LISV-GEESFNLKOKVTFFHVLISVAVSCSSLKEHNF--PFAI-----RVFSLLQK 735  
 DB 1237 LQTIQDLETSLNLOMQSOEI--SGLKQCEIDAEKYYISGPHLSTSQDNNAHLQCSLQ 1294  
 QY 736 KIKKLESVITAVEISEMHIEMLDRGIPVELMAHYVEELNSTQVAVADSVFLFSLKX 795  
 DB 1295 TNNKLNLEKICEIILQAEKYL-----VELNDSRECECTATRKMAEBYK 1340  
 QY 796 FIYALK-----APKSPKADIWNPEOLK-----EDSRDYLHLIGLFEMM 836  
 DB 1341 LANEVKIINDSGLHGLVEIDIPGEPGEQNEQHPVSLAPLDESNSYEHITLSDKEVQ 1400  
 QY 837 LMGADAVHFRVIMKLFIKYHLEDVF-----OLFRCVLYMTYGSLS-----SNPLN 882  
 DB 1401 M-----HFAELQERFLSLOSEHKITLHDQCCWSSKMSLQTYVLSLKAENVLSTNLN 1454  
 QY 883 CSVKTVLOQALVYVCGAMLSQKOCKHOLASIS--SPVVTSLINLNGSPVKEVRRAA 938  
 DB 1455 PGQDLVKEMQ-LGEBGLVPSLSSGCVDPSSLSLGGSSFRLALBOGD----- 1504  
 QY 939 IOCLQALSGVAPFVLIIDHLISKA-----EITSDAVVIDLATLFELORE-KK 989  
 DB 1505 MSLSLNLEGAVASANOCYDEVFCSSIOEENLTKETPSAPAKGVELESELCEVYQSLK 1564  
 QY 990 LKSHQKLETLKMLSCVYSCPSYIAKDLMKVLOQNGE-----VYLSQLLMAEBOLK 1044  
 DB 1565 LEKKNESQIGIMKN-----KEIOLEQLUSSEROELDCLRKQYLSENEQWOK 1611  
 QY 1045 I-----QKEPTAVLKDAMVLIHLTLGKYNEFSVSLNEDP-----KSLD 1083  
 DB 1612 LNSVTLKESKLAERKQTOBLSLEVARLOLOGJLDSRRLIGITDREDAIQGRNESC 1671  
 QY 1084 IFIKAVHTTKELVAGMPTIQTALKEKITKPPFAAISDEKVOQKLLRMLFDLLVNCNSHC 1143  
 DB 1672 --ISKHEHSE-----TTERTPGHADVHOICDXAOODL-----NNDIEXI 1708  
 QY 1144 AQTSSVVFQISVANNEQVRIELEPPDKAKPLGTVQKRRQKQKKSQDLESVOEGSGY 1203  
 DB 1709 TETGAKPTGECGSGOSPNTYEPGEDTQGS-----SECISELSTFS- 1751  
 QY 1204 WQRTVLTLELLOHKKLSRQILVFTLFN-----LH-----SRCEPLPQEOGNE--Y 1250  
 DB 1752 GPNALVPMQFLONGEDIHNLQARVETSENURLHVEDRBRKYESLLENKEKELDSKLH 1811  
 QY 1251 TKQILISCLLNTCQKLSPDGKIIPKDIIDEEKFNVELIYQCIIRLSEMPOTHHALLLOT 1310  
 DB 1812 LOEVOJMTKEIACIELEKIVGELKKNESDLS-----KLEBFSCDHQHELLQRYET 1861  
 QY 1311 VAGIPDKVLAHNMISFTFMGANVRLDPTYSFOVYINKTVKAVIPALLOSDEG-----DSI 1366  
 DB 1862 SEGLNSDLEMHADKSRDIGNVAKANDSKERFLDYVENEL--SRIRSEKASTIEHML 1918  
 QY 1367 EVSRNVEEIV-----KIISVFVALPHVPEHRLPILVQL--VDTLGAEK 1410